FACULTEIT ECONOMIE EN
BEDRIJFSWETENSCHAPPEN

KU LEUVEN

New perspectives on mutual fund decision analytics: Contributions from market efficiency and social responsibility

Proefschrift voorgedragen tot
het behalen van de graad van
Doctor in de Toegepaste
Economische Wetenschappen aan
de KU Leuven en Vrije Universiteit
Brussel

door

Tim VERHEYDEN

NUMMER 472 2015

Committee

Advisors	
Prof. dr. Lieven De Moor	Vrije Universiteit Brussel
	KU Leuven
Prof. dr. Geert Van Campenhout	KU Leuven
Co-advisor	
Prof. dr. Luc Van Liedekerke	Universiteit Antwerpen
	KU Leuven
Internal committee members	
Prof. dr. Rosanne Vanpée	KU Leuven
Prof. dr. Kris Boudt	Vrije Universiteit Brussel
	Vrije Universiteit Amsterdam
External committee member	
Prof. dr. Rob Bauer	Maastricht University

Acknowledgements

My PhD journey has been nothing but an absolute pleasure. I now realize that this was only possible because of the many great people that have accompanied me on this trip.

First and foremost, I am forever indebted to Prof. Lieven De Moor. Lieven was the one who offered me a PhD scholarship on August 28, 2012, my birthday and the day before I left for one of my great foreign adventures. Even though I still had another year left as a master student, Lieven somehow had faith that I would do well in the future. He is also the one who always allowed me to work in my own way, not forcing me to waste time in an often inefficient work environment. He always approved of my numerous travels and the additional challenges I took on outside of academia, even though they often interfered with my doctoral work. And he inspired me to dream big and only aim for the very best, be it inside or outside of academia. If I managed to finish my PhD in 1.5 years, it is only because Lieven was there to guide and mentor me along the way.

I am also grateful to the members of my doctoral committee, who approved of my unorthodox way to finish my PhD. Prof. Luc Van Liedekerke has been an inspiration from day one, and a great source of knowledge for the social responsibility part of this PhD. For helpful and insightful comments, I could always count on Prof. Geert Van Campenhout. Prof. Rosanne Vanpée was an absolute pleasure to work with on the third chapter of this dissertation. I would also like to thank Prof. Kris Boudt, who often challenged me with critical remarks on the doctoral seminars. Finally, I owe a debt of gratitude to Prof. Rob Bauer, with whom I had some very pleasant conversations on the role of sustainability in finance. He also directed me to Arabesque, which led to me landing what seems to be the best job in the world.

My thankfulness also goes to a number of friends, to only some of whom it is possible to give particular mention here. The fact that I have always enjoyed my life as a PhD student is largely due to Capitant. Through my experience at Capitant, I had the pleasure to meet a great number of interesting and genuinely nice people. My Capitant Belgium colleagues, Thomas, Shana and Michaël, have become great friends and I am eternally grateful for the fact that they allowed me to work from abroad. I also want to thank my former team at Capitant Leuven, who made my life in Leuven a lot more enjoyable. Along the way, I also had the pleasure to get to know Wouter and Laetitia, two wonderfully interesting individuals that always make my day. In general, I am grateful for the unique Capitant atmosphere, created by everyone involved with Capitant, as it always helped me to recover from a hard day of PhD work.

Next to my Capitant buddies, some other friends also accompanied me along the way. I want to devote particular attention to my so-called MCDA family that I met in Hamburg whilst on academic duty. Renaud, Raffaele, Gabriella, Mehmet, Audrey, Melis and Şahan, thanks again for the wonderful times and I very much look forward to meeting you somewhere around the world in the future. I am also thankful for the wonderful people I met along my journeys in Norway, Thailand, Lisbon, London, Rethymno and Peru. Anne-Sofie, I also want to thank you for your friendship. Even though we don't get to spend all that much time together, your like-minded spirit always gave me a boost to continue working and enjoying life in general. Finally, I am forever indebted to Elisabeth, who I now only know from my past. Thank you for helping me become who I am today.

My last words of gratitude go to my wonderfully complex family. Throughout my life, my father has never stopped believing in me, even at times when I didn't believe in myself. He also made my life as a student incredibly easy. Mom, you have been my emotional safe place; always there when I most needed you. Sis, you will forever be my comedic soulmate, my beacon of genuinely bizarre and entertaining philosophical discussions. Even though I sometimes fail to understand the messed-up structure of this large family, I wouldn't trade you for the world as I have always felt at home.

Summary

Throughout the 21st century, investors have seen interest rates declining. Most recently, rates have gone down in response to the global financial crisis that was triggered by the collapse of the U.S. housing market. By lowering interest rates, central bankers aim to support the global financial system and spur consumption and investments. Not surprisingly, however, this policy was not immediately successful, as investors had taken a hit from the market collapse and the average consumer was frightened by the negative reporting and the poor economic outlook.

Only half a decade after the start of the most recent financial collapse, investors and consumers are regaining confidence as more and more signs are pointing towards a mild economic recovery, especially in the U.S. market. Consequently, investors are becoming increasingly displeased with returns on traditional savings products, which have attracted a great amount of post-crisis capital despite declining interest rates. Alternatively, money is now flowing away from savings accounts in a search for higher yield.

Considering investor's increased caution due to the recent past, and the negative environment for fixed income investments, one of the most appealing alternatives to a savings account are mutual funds. A mutual fund provides a low-threshold access to financial markets and its higher returns, whilst avoiding the need for much active management or diversification effort on the side of the investor. However, the decision to invest in a mutual

fund immediately triggers a much more difficult question: "what mutual fund to invest in?"

With a supply of tens of thousands of mutual funds globally, picking a mutual fund can be just as hard as picking an individual stock. The most typical approach is to consider criteria like return, size, expense ratio, sharpe ratio and style. Additionally, some third party research providers also give out fund ratings, which are increasingly used in mutual fund selection decisions. Nevertheless, a combination of the above criteria rarely leads to a definitive insight on what fund to pick, but merely helps the investor in making educated guesses on what fund will perform best in the future.

To further help improve the investor's mutual fund decision process, this PhD thesis provides a new perspective from two different angles: market efficiency and social responsibility. In the first part, we consider the financial calamities from the recent past by looking into market efficiency. Doing so, we enable investors to better understand the performance profile of funds and the way fund managers deal with periods of market distress. In part two, we devote serious attention to the emerging trend of social responsibility in investing, by assessing how fund managers incorporate the concept of sustainability into the investment process. This can help investors to better understand the sustainable nature of a fund and how the fund manager is incorporating social responsibility criteria to drive financial performance. The market efficiency and social responsibility perspective are also linked through their common impact on downside risk.

Part I — Market Efficiency The concept of market efficiency has been subject to debate for over half a century. In that time, two opposing views were developed: the efficient market hypothesis (EMH) and behavioral finance. Both represent the extreme ends of a spectrum of beliefs on market efficiency. The joint awarding of the Nobel Prize in Economics to proponents from both sides of the debate (resp. Eugene Fama and Robert Shiller) confirmed the academic stalemate.

Following the construct of time-varying market efficiency (Campbell, Lo, & MacKinlay, 1997), Lo (2004, 2005) developed a reconciliating framework called the adaptive markets hypothesis (AMH). Under the AMH, markets can be relatively efficient for a long time, until an external shock causes the market ecology to change, triggering a time of relative market inefficiency. Once a new market ecology is formed, market efficiency is restored and the market returns to equilibrium.

The first empirical evidence was supporting the idea of time-varying market efficiency, but pointed towards an inverted pattern of efficiency. However, the development of new tests, exploiting different proxies for weak form market efficiency, allowed researchers to further look into the AMH. Implementing six state-of-the-art rolling weak form market efficiency tests, we further confirm the concept of time-varying efficiency. Furthermore, we are able to confirm the predicted pattern of efficiency from the AMH. Finally, we find the impact of the most recent financial crisis to be most prominent on the U.S. stock market, while disruptions in weak form market efficiency on the European and Japanese markets are more rare in the last fifteen years.

Given the empirical support for the AMH, we also look into the ability of fund managers to anticipate market efficiency and exploit market inefficiencies. Combining a proxy measure for time-varying efficiency and performance, we find a positive relationship between weak form market efficiency and α. The majority of funds appear unable to systematically outperform the market, although the market efficiency perspective seems to help in detecting a number of top-performing funds. A good fund manager can be distinguished from his or her ability to limit drawdown and downside risk in times of relative market inefficiency, whilst still fully reaping the benefits when markets return to equilibrium. Conditioning mutual fund performance on market efficiency, we construct a so-called conditional alpha ratio, which helps identify top-performing funds that might be of interest to a mutual fund investor.

Part II — Social Responsibility Socially responsible investing (SRI) has moved from a niche to a more mainstream investment strategy over the last decade. Originally developed from religious principles and negative screening, SRI strategies now also include positive screening and active engagement. In the strategic management literature, there is compelling evidence of a positive relationship between corporate responsibility and financial performance. In financial economics, most research finds no significant performance differential between SRI and conventional funds.

Several challenges are facing the further development of SRI. Because of the increasing supply of SRI products, investor's have a hard time comparing and understanding different SRI offerings. This calls for the development of a SRI decision tool that gives insight into the way fund managers incorporate sustainability in the investment process. Such a tool can also help grow the retail end of the SRI mutual fund market, which is currently relatively underdeveloped compared to the overall mutual fund market.

We consider multi-criteria decision analysis (MCDA) as the methodology to develop such a SRI decision tool. We find and implement four specific methods to construct a process-oriented SRI score for mutual funds, focusing on the process used by fund managers to translate social responsibility criteria into the mutual fund investments. From an extensive robustness analysis, we find the PROMETHEE outranking approach to be most suited as a method for the SRI decision tool. The PROMETHEE net flow scores can be used by a mutual fund investor to help understand and compare different (SRI) funds.

One problem with the PROMETHEE-scores is the lack of a parametric way to assess scoring differentials. Therefore, we implement a MCDA sorting tool called FLOWSORT. Using FLOWSORT, we can transform the PROMETHEE-scores into different categories using reference profiles. The sorting of funds from the underlying PROMETHEE-scores serves as a non-parametric way of assessing scoring differentials and further adds to the decision process of mutual fund investors when considering social responsibility. We also illustrate how the FLOWSORT groups can be used to improve the performance research on SRI funds. Typically, researchers only consider two samples of funds: SRI vs. conventional funds. The problem with this approach is that the definition of the fund is biased as it comes from the fund manager and that the idea of sustainability is reduced to an all-or-nothing concept. From the broad definition of process-oriented SRI and the different groups of FLOWSORT, we overcome these two issues and confirm that there are no performance differentials between any of the groups for the Belgian mutual fund market. Upon availability of appropriate data for other mutual fund markets, this approach can be replicated to strengthen earlier performance research.

The various chapters in this thesis can be found in:

(i) Verheyden, T., Van den Bossche, F., and De Moor, L. (2013). A tale of market efficiency. *Review of Business and Economic Literature*, 58(2):139–156.

(ii) Verheyden, T., Van den Bossche, F., and De Moor, L. (2014). Towards a new framework on efficient markets. Working paper.

(iii) Verheyden, T., De Moor, L., and Vanpée, R. (2014). Mutual fund performance: A market efficiency perspective. Working paper.

(iv) Verheyden, T. and De Moor, L. (2014). Multi-criteria decision analysis: Methods to define and evaluate socially responsible investments. *International Journal of Management and Decision Making*, forthcoming.

(v) Verheyden, T. and De Moor, L. (2014). Process-oriented social responsibility indicator for mutual funds: A multi-criteria decision analysis approach. Working paper.

(vi) Verheyden, T. and De Moor, L. (2014). Sorting mutual funds with respect to process-oriented social responsibility: A FLOWSORT application. *Decision Science Letters*, 3(4):551–562.

Table of contents

Part I

Market efficiency

Chapter 1

A tale of market efficiency

Abstract

The efficient market hypothesis (EMH) has been subject to debate for decades. The field of behavioral finance was developed in response to the body of anomalous evidence with regard to the EMH. Considering theoretical and empirical research, we summarize the debate on weak form market efficiency. Testing methodologies developed in the early aftermath of the first discussions are explored and recent alternative approaches are reviewed. As a way out of the stalemate, we consider Lo's adaptive markets hypothesis (AMH), which yet has to enter mainstream academic finance. From our review of the tale of market efficiency, we suggest that future research is geared towards the further development of time-varying test methodologies and the corresponding theoretical framework.

1.1 Introduction

Driven by the desire to understand and explain the impact of investors' decisions on financial markets, the debate on the informational efficiency of stock markets has been going on now for more than 50 years. In a larger historical perspective, one could go back to the 18^{th} century to see that even Adam Smith (1759, 1766) was troubled by the efficiency and self-stabilizing nature of financial and economic markets, which essentially boils down to the question whether or not stock prices are in line with the intrinsic value of the underlying financial asset.

Fama (1970, p. 383) defines an efficient market as "a market in which prices always fully reflect available information" and makes a distinction between different types of efficient markets based on three concretions of the concept "available information" i.e. weak form efficient markets (historical price information); semi-strong form efficient markets (all publicly available information); and strong form efficient markets (all information, both public and private). In weak form efficient markets it is impossible to persistently generate portfolio returns higher than the market return by trading on past price information i.e. technical analysis of stocks is obsolete[1]. A semi-strong efficient market implies that it is impossible to persistently beat the market by using a trading strategy based on public information (e.g. newspapers) i.e. fundamental analysis is ineffective[2]. If markets are strong form efficient, even insider trading on private information will not be able to outperform the market portfolio, besides by pure luck. This paper focusses on the weak form of the EMH bearing in mind that rejecting the null hypothesis of weak form efficient markets naturally

[1] Technical analysis consists of investigating time series of past prices and returns of a stock in order to derive a certain pattern that can be extrapolated to make profitable predictions of future price movements (Brown & Jennings, 1989).

[2] Fundamental analysis consists of analyzing all publicly available information (e.g. financial statements) about a certain stock to infer important insights that can be used to make a profit in the stock market (Kothari, 2001).

leads to the rejection of the semi-strong and strong form of the EMH.

Proponents of the EMH argue that if the price of a stock would appear to be too high given past price information, rational investors would bid the price down to make a profit and vice versa. What they call "the wisdom of the crowds" would eventually force stock markets to be efficient (Fox, 2009). More generally, they believe that investors are rational optimizers that are able to make the best possible decisions given certain information (e.g. past price information). Proponents of behavioral finance, on the other hand, believe that investors are not always fully rational and therefore are not able to force the stock market to be efficient at all times (e.g. Shefrin, 2000). They refer to recent bubbles and financial crises to point out that there are different psychological effects that cause human beings to stray from rational decision making.

There is still no consensus on the validity of the EMH. Nevertheless, valid financial models are important for policy makers and investors. One example is the well-known theory of diversification deducted from the optimal portfolio theory (Markowitz, 1952). However, one needs to stay critical even with respect to well-established theories, as the world is not a static environment. When academic theory is flawed it has the potential to set the entire economy astray (Fox, 2009; Nocera, 2009). One example is the housing bubble that caused the 2008 financial crisis. While policy makers, banks and investors were blindly following the bullish market, irrational exuberance was building up underneath (Shiller, 2000). Many years later, we are still trying to deal with the consequences, and even the future of an entire generation is at stake.

The validity of financial theories can be tested by developing appropriate methodology. For the EMH, methodologies using different proxies for the concept of weak form market efficiency were established. Earlier research on efficient markets signals that there might be some flaws to these test methodologies (Campbell, Lo, & MacKinlay, 1997; Lim & Brooks, 2011), leading to the preservation of the wedge driven between opponents and proponents of the EMH.

In this paper we provide an overview of the theoretical and empirical debate on market efficiency. Additionally, we review the development of statistical tests for weak form market efficiency and a promising theoretical framework that might reconcile the opposing views of the EMH and behavioral finance. From our overview we find that further research is needed on the link between the theoretical and empirical pattern of efficiency.

Other excellent reviews of the market efficiency debate already exist. Sewell (2011) provides an in-depth overview of the historical developments leading to the establishment of the EMH. Lim and Brooks (2011) give a more complete overview of the testing methodologies for the different versions of the EMH. Dimson and Mussavian (1998) present an overview of the developments in market efficiency research and observed anomalies. Our current paper contributes to the market efficiency debate by combining a concise overview of the theoretical and empirical debate with the developments of weak form market efficiency test methodologies.

1.2 Theoretical debate on market efficiency

Most of the theoretical advancements in the development of a market efficiency theory were made during the 1960s and 1970s. However, some earlier contributions were crucial in the later establishment of the EMH. Louis Bachelier's (1900) "Théorie de la spéculation", rediscovered by Leonard Savage and Paul Samuelson in 1955, modeled the stochastic process of the Brownian motion. His model was the mathematical foundation for Paul Samuelson's (1965) martingale theory that became one of the centerpieces of finance and the research on the efficiency of financial markets. The first statement about the efficiency of financial markets came from George Gibson (1889, p. 11): "When shares become publicly known in an open market, the value which they acquire may be regarded as the judgment of the best intelligence concerning them.” Alfred Cowles (1933, 1944) and Holbrook Working (1934), founding members of the Econometric Society, laid out the foundations for an informationally efficient stock market by showing that investors are unable to beat the market by means of price forecasts and that stock returns exhibit behavior similar to lottery numbers.

The actual establishment of a theory on efficient markets started with Samuelson (1965), who theoretically proved that in an informationally efficient stock market, where the stock price contains all available information and expectations from market participants, prices fluctuate randomly. Roberts (1967) was the first to coin the term efficient market hypothesis (EMH) and suggested a distinction between several types of efficiency.

Fama (1970) defined an efficient market as a market that fully reflects all available information. He introduced three types of informational efficiency: weak form, semi-strong form and strong form efficiency. Summarizing empirical results from weak form, semi-strong form and strong form efficiency tests, Fama concluded that financial markets are efficient in at least the weak sense. Although some price dependencies were found, they never sufficed to be used in profitable trading mechanisms. Fama (1970) also pointed out a joint-hypothesis problem. Essentially, it is argued that it is impossible to ever correctly test the efficient market hypothesis, because no academic consensus is found on the true underlying asset-pricing model. Whenever a test of market efficiency would reject the efficiency hypothesis, there is always the possibility that it is simply due to the underlying asset pricing model finding an incorrect theoretical asset value. The only conclusion that can be made from efficiency tests is that a market is efficient or not with respect to a certain underlying asset pricing model. Alternative definitions of market efficiency were proposed by, for example, Jensen (1978, p. 96): “A market is efficient with respect to information set θ_t if it is impossible to make economic profits by trading on the basis of information set θ_t”; or Malkiel (1992) who stated that a stock market is efficient whenever the prices of stocks remain unchanged, despite information being revealed to each and every market participant.

The theoretical requirements for market efficiency rest on three underlying levels of development (Ackert & Deaves, 2010; Shleifer, 2000). The first level contends that all investors are rational at all times, which would automatically lead to markets being efficient. If this were to be untrue, a second level of support argues that irrationality is uncorrelated across investors, which would lead to the noise trading being canceled out. This would render markets efficient after all. The ultimate level of defense used by advocates of market efficiency is that even if irrationality is correlated among investors, some smart money traders can benefit from arbitrage to bring market values back in line with underlying fundamentals. The assumption here is that there are no limits to the arbitrage pricing mechanism.

An important theoretical critique of the EMH was formulated by Grossman and Stiglitz (1980), following the earlier work of Grossman (1976). They argue that in order for investors to be motivated to spend resources for collecting and analyzing information to trade on, they must have some form of incentive. If a stock market would prove to be perfectly efficient, however, there would be no reward for collecting information, since that information would already be reflected in the current stock price. This simple paradox shows that financial markets can never become perfectly efficient, since no investor would be motivated to collect information in the first place. Consequently, no one would trade on new information and it would become impossible for stock market prices to reflect all available information.

In response to some of the early critiques on the EMH, two fundamental levels of criticism were formulated from a behavioral finance perspective, which merges the traditional theory of finance with concepts from other social sciences like psychology and sociology. On the one hand, behavioral scholars documented multiple behavioral biases that directly contradict with the first two levels in support of the EMH. Kahneman and Tversky (1979) argued that people tend to be loss averse as they hate losing more than they love winning. They incorporate this behavioral bias in their prospect theory and suggest that, rather than simply multiplying probabilities and utility as in von Neumann and Morgenstern's expected utility theory (1944), losses and gains should be treated differently. Barberis, Shleifer and Vishny (1998) explain the short- and medium-horizon underreaction to information (or the momentum effect) using the concept of conservatism i.e. investors erroneously believe that the earnings process underlying stock prices is mean-reverting and so they underreact to news. To explain the longer-horizon overreaction, they refer to the representativeness heuristic i.e. investors overextrapolate from a sequence of growing earnings, and therefore overreact to a long trend. Daniel, Hirshleifer and Subrahmanyam (1998) related overreaction to overconfidence, as traders tend to overestimate the precision of their private signals, leading to prices being pushed above the fundamental level in the case of good news.

On the other hand, behaviorists have criticized the ultimate level in support of the EMH, i.e. that there are no limits to arbitrage. A seminal paper to this regard is Shleifer and Vishny (1997). They show that arbitrageurs are mostly highly specialized investors who are limited in their ability to benefit from pricing anomalies under extreme circumstances like high volatility and low liquidity. Two types of risk present themselves. There is the risk that new information arriving in the market calls for a reevaluation of the stocks addressed by arbitrage, i.e. fundamental risk. Additionally, Shleifer and Vishny (1997) point out noise-trader risk using the example of two German bond futures contracts. In the short run, mispricing can get worse before it gets better because of new noise traders entering the market. This observation is in line with the famously quoted John Maynard Keynes: "Markets can remain irrational longer than you can remain solvent" (Lowenstein, 2000, p. 123). Next to these two types of risk, Gromb and Vayanos (2010) find that several implementation costs further limit the effectiveness of arbitrage.

Following most of the theoretical underpinnings of the EMH, and the behavioral critiques, several empirical studies were conducted to look for validation. An overview of the empirical response to the theoretical debate is presented in the next section.

1.3 Empirical debate on market efficiency

Empirical testing of the EMH was greatly accommodated by the development of asset-pricing models, which are based on the notion of informational efficiency. Building on the earlier work of Markowitz (1952), Sharpe (1964) developed the capital asset pricing model (CAPM) that allows for the calculation of a theoretical rate of return on an asset, given the amount of non-diversifiable risk the asset entails. The reason for taking only non-diversifiable risk into account is the assumption that the asset is added to a well-diversified portfolio that neutralizes idiosyncratic risk to all extent. However, in later years, scholars came across some interesting asset pricing anomalies and argued that the CAPM was too limited by accounting for only one factor of risk. Stephen Ross (1976) came up with an alternative: the arbitrage pricing theory (APT). APT is far more flexible and states that the expected return on an asset is a linear function of different factors of risk, each with their respective factor sensitivity. Whenever the actual return on the asset deviates from the one derived from the theoretical model, the force of arbitrage will bring the actual rate of return back in line with the theoretical one. However, the APT is very general and does not give any guidelines as to what specific factors of risk to account for. Fama and French (1993) extend the CAPM to a three-factor model. Starting from their observation of pricing anomalies with respect to market capitalization and growth versus value strategies, Fama and French (1993) found the expected rate of return to depend on the exposure of the asset to each of three factors: market risk premium (non-diversifiable risk), market capitalization (size) and the book-to-market ratio (valuation). Following the momentum puzzle pointed out by Jegadeesh and Titman (1993), Carhart (1997) extended Fama and French's three-factor model to a four-factor model, taking into account a momentum risk factor. Asset pricing theory is very important in the debate on efficient markets since it provides researchers with the ability to theoretically derive the price of financial assets. That way, it is possible to examine whether actual returns on assets are in line with the theoretical rate of return derived from an underlying asset pricing model. However, it is never entirely certain what the correct theoretical price of an asset is, so a joint-hypothesis problem (Fama, 1970) arises.

In one of the earliest empirical studies on efficient markets, Fama (1965a) empirically showed that financial markets follow a random walk. Additionally, technical and fundamental analysis cannot possibly yield risk-adjusted excess returns (Fama, 1965b). Fama and Blume (1966) pointed out that no economic profits could be made using technical trading rules, like the filter rule in Alexander (1961, 1964), as trading costs are too high[3]. An excellent overview of most empirical work in support of the EMH can be found in Fama (1970, 1991, 1998).

The empirical work supporting the behavioral critique of investor irrationality goes back to De Bondt and Thaler (1985), who tested the hypothesis that investors tend to overreact to unexpected and dramatic news events. They constructed portfolios based on market-adjusted cumulative abnormal returns over three years: the top 35 stocks were allocated to a winner portfolio and the bottom 35 stocks were allocated to a loser portfolio. Then they looked at the performance of the portfolios over the subsequent three years. They found that the loser portfolio had outperformed the winner portfolio by, on average, 23%, which is consistent with

[3] An example of an x% filter rule: buy and hold securities of which the daily closing price moves up by at least x%, until the price moves down by at least x% from the subsequent high, at what point it is time to simultaneously sell the security and go short. The short position is then maintained until the daily closing price of the security rises at least x% above the subsequent low, after which the short position needs to be covered and the security is bought again. A very small-width filter is a filter in which x lies between 0.5 and 1.5 (Alexander, 1961, 1964).

the overreaction hypothesis to new information. Lakonishok, Shleifer and Vishny (1994) conducted similar empirical research using proxies for value instead of historical price information. Delong, Shleifer, Summers, and Waldmann (1990) showed that the long-horizon negative autocorrelation in returns (or the reversal effect) can be explained by a stylized model with two types of agents: fundamentalists, who get signals about intrinsic values, and chartists, who learn indirectly about intrinsic values by looking at prices. Whenever a good signal is received by fundamentalists, prices will increase. Chartists will observe this rise in prices causing some of them to buy, which in turn further increases prices and causes more chartists to buy. Eventually, share prices are so far beyond intrinsic values that fundamentalists start selling again. An excellent overview of the empirical work on market anomalies can also be found in Keim and Ziemba (2000) and Schwert (2003).

Several empirical studies following Shleifer and Vishny's (1997) finding of limits to arbitrage were conducted as well. For example, Mitchell, Pulvino and Stafford (2002) study a number of arbitrage opportunities. What they find is that, even though there is an obvious mispricing in the value of parent companies with respect to their subsidiaries, it can take a significant amount of time for an arbitrage strategy to become effective. Significant evidence of fundamental and noise trader risk, and implementation costs were found, which confirms the theoretical arguments of limits to arbitrage. More recently, limits to arbitrage were examined in the light of the 2007-2008 financial crisis. From a large set of international cross-sectional data, for example, Buraschi, Sener and Meguturk (2012) find that the most important explanation of limits to arbitrage is the interaction between leverage constraints and funding costs.

1.4 Weak form market efficiency tests

Following our summary of both the theoretical and empirical debate on market efficiency, we now provide an overview of the development of traditional and alternative tests for weak form market efficiency.

Together with the valuable data associated with anomaly events like the 1972 Black Monday crash, the evolution of computing power allowed researchers to come up with new and more formal empirical tests of market efficiency (Bodie, Kane, & Marcus, 2010). We focus here on three particular types of statistical tests of weak form market efficiency that were developed in the early aftermath of the EMH.

The first group consists of tests that are based on the return autocorrelation proxy of efficiency. The general philosophy behind these tests is the following: if significant autocorrelation is found among the returns, there is some extent of predictability in return time series, which is in contradiction with the efficient market hypothesis. Return autocorrelation is also very much related to the concept of technical analysis in which investors believe that historical price series exhibit regularities that can be profitably exploited to extrapolate future price movements. The test of autocorrelation is therefore also a test for the applicability of technical analysis. In general, empirical investigations have led to the conclusions that autocorrelations in short horizon returns (day, week, month) tend to be positive for returns on portfolios and negative for returns on individual stocks; autocorrelations in medium horizon returns (1-12 months) tend to be positive; and the long horizon (1-5 years) return autocorrelations tend to be negative.

Granger and Morgenstern (1963) and Fama (1965a) use more formal linear serial correlation tests to examine market efficiency. Lo and MacKinlay's (1988) variance ratio (VR) test is of the same kind. The VR is the ratio of the k-period return variance over k times the variance of the one-period return. If stock prices are following a random walk their k-period return variance should be the same as k times their one-period return variance. So, it suffices to test whether or not the VR is significantly different from one to test the random walk theory and therefore the informational efficiency of stock prices. Formally, the test looks the following:

$$VR(k) \equiv \frac{Var(r_t(k))}{kVar(r_t)} = 1 + 2\sum_{j=1}^{k-1}(1 - \frac{j}{k})\rho_j \tag{1.1}$$

With $r_t(k) \equiv r_t + r_{t-1} + \cdots + r_{t-k+1}$ and ρ_j being the j^{th} order autocorrelation coefficient of r_t. $VR(k)$ is a particular linear combination with linearly declining weights of the first $k-1$ autocorrelation coefficients of r_t. The following orders of differentiation k have been suggested by Lo and MacKinlay (1988): 2, 4, 8 and 16. From their test, Lo and MacKinlay (1988) find that the random walk hypothesis does not hold for weekly stock market returns. An excellent overview of recent developments in variance ratio testing is provided by Charles and Darné (2009).

Unit root tests examine the weak form efficiency of stock returns in an alternative way. The basic idea is that stock returns that contain a unit root, i.e. are non-stationary, follow a random walk. An example is the augmented Dickey-Fuller (ADF) test executed on log returns. In later years, innovations were added to the ADF test to examine weak form market efficiency (e.g. Narayan & Smith, 2007). However, Rahman and Saadi (2008) have shown that the existence of a unit root in stock returns is not a sufficient prerequisite for the random walk hypothesis to hold. In addition to stationarity, return series need to be serially uncorrelated.

Granger and Andersen (1978) pointed out that stock markets could exhibit inefficient behavior even if the linear autocorrelation tests are in line with informational efficiency. This conclusion triggered the development of a third type of weak form efficiency tests based on non-linear serial dependence. Examples are the Hinich bicorrelation test (Hinich, 1996), the Engle LM test (Engle, 1982) and the BDS test (Brock, Scheinkman, Dechert, & LeBaron, 1996). These tests are also included in the convenient nonlinearity toolkit developed by Patterson and Ashley (2000).

Inspired by the seemingly irreconcilable opinions in the efficient markets debate, Campbell et al. (1997) suggested a new approach in which the degree of market efficiency is tested over time resulting in more nuanced conclusions. This idea triggered the development of alternative tests of market efficiency that capture efficiency in a more dynamic way than the tests developed shortly after the EMH was established.

A first alternative testing approach looks at separate time windows and the evolution between those windows. This non-overlapping sub-period analysis is useful for examining the impact of a specific policy reform (e.g. short sell prohibition) on market efficiency. Obviously, these tests look at the changes in efficiency rather than trying to proof that stock markets are weak-form efficient or not over a fixed period of time. Examples of this approach can be found in the work of Kim and Singal (2000a, 2000b), Jain (2005) and Lim (2008).

A second alternative test transforms a data sample of n observations into $n - w + 1$ overlapping windows of width w, which accommodates the measurement of gradual changes in market efficiency irrespective of specific policy reforms[4]. Examples of this rolling estimation windows procedure are: rolling variance ratio tests (Tabak, 2003; Kim and Shamsuddin, 2008), rolling augmented Dickey-Fuller unit root tests (Phengpis, 2006), rolling bicorrelation tests (Lim, 2007), rolling parameters of ARCH models (Alagidede & Panagiotidis, 2009) and rolling Hurst exponents (Costa & Vasconcelos, 2003). A drawback of these rolling versions of static tests is the issue of the robustness of results with respect to the width of the time window.

Third, the time-varying parameter approach uses state space models that

[4] For example, for a time series of 100 observations and a window width of 20, the time series data is transformed into 81 overlapping windows with a width of 20. The first window goes from observation 1 to 20, the second from 2 to 21 and so on.

enable regression parameters to change over time[5]. In this way, regression methods can be extended to measure time-varying efficiency in emerging and developed stock markets. For example, Zalewska-Mitura and Hall (1999) use a GARCH-M approach to model financial day-to-day data on emerging markets and let the parameters evolve through time by applying a Kalman filter[6]. Basically this approach generates a time-varying autoregressive parameter of the return variable. Kim, Shamsuddin, and Lim (2011) also apply this test to developed stock markets.

1.5 Reconciliating framework

From all of the empirical studies using the variety of available weak form market efficiency tests, Fama (1998) concludes that there is a lack of valid empirical evidence to disprove the EMH. However, advocates of behavioral finance did not rest their case either (e.g. Thaler, 1992; Shleifer, 2000; Shiller, 2000; and Shefrin, 2000). Shiller (2003) claims that the philosophy of efficient markets should remain a characterization of an ideal world but is not an accurate description of global financial markets. In the same journal, literally preceding Shiller's article, Malkiel (2003, p. 80) argues: "If any $100 bills are lying around the stock exchanges of the world, they will not be there for long." His statement became a classical economic joke to explain that efficiency anomalies would not persist because someone

[5] "State space modeling provides a unified methodology for treating a wide range of problems in time series analysis. In this approach it is assumed that the development over time of the system under study is determined by an unobserved series of vectors $\alpha_1, \cdots, \alpha_n$, with which are associated a series of observations $y_1, \cdots, y_n$; the relation between the α_t's and the y_t's is specified by the state space model. The purpose of state space analysis is to infer the relevant properties of the α_t's from a knowledge of the observations $y_1, \cdots, y_n$." (Durbin & Koopman, 2008, p. 1)

[6] The GARCH-M approach is a way of dealing with heteroscedastic and autocorrelated errors taking into account the financial risk premium property. The Kalman filter is a statistical algorithm that uses data in a step-by-step way by pushing time windows forward.

will immediately benefit from the opportunity through the price arbitrage mechanism.

Lo (2004, 2005) tries to reconcile the EMH and behavioral finance by applying evolutionary biology. Starting from the concepts of bounded rationality and satisficing[7] (Simon, 1955) and the notion of biological evolution, Lo's adaptive markets hypothesis (AMH) states that many of the behavioral biases, found in behavioral finance, follow a certain evolutionary path and the degree of financial market efficiency depends on the strength of this underlying evolutionary force. As investors act in their own self-interest but human rationality is bounded, investors will make mistakes. If human beings are reluctant to learn from their mistakes, markets will be more likely to exhibit higher levels of inefficiency. However, if investors quickly learn from mistakes and adapt to new market conditions, temporary levels of inefficiency will only survive for short periods of time. This learning and adaptation process will be driven by competition among investors, and natural selection will decide which investors are driven out of the market and which investors can stay. This natural selection process shapes the new market ecology and its evolutionary dynamics. As long as there is no shock that causes market ecology to change, stock markets are fairly efficient. However, once a certain event triggers the process of competition and natural selection, markets become temporarily less efficient. Once the new market ecology is formed, efficiency of financial markets returns to pre-shock levels.

Lo's AMH theory reconciles the EMH and behavioral finance by stating that markets are not perfectly efficient all the time, nor are they inefficient. There is a certain evolutionary aspect to the process of market efficiency. For a long time, stock markets can process information in a reasonably

[7] Both concepts explain that humans at times behave in a less rational way, hence their rationality is bounded. Furthermore, humans do not have the information, nor the methodology to always optimize in a rational way. Consequently, humans will use some rules of thumb or heuristics to find satisfactory results that are not necessarily completely rational i.e. satisficing.

efficient manner, and the EMH seems to apply. However, a certain shock, crash or other event might disrupt this state of efficiency. Some market participants are driven out of the market and some new participants enter the market. During this process in which a new market ecology is formed and participants learn from and adapt to new market conditions, relative levels of inefficiency are found. Once the transformation period ends, levels of market efficiency are restored, until a new crash, shock or other event disrupts the ecological equilibrium.

Looking at the 2008 financial crisis we indeed recognize elements from Lo's theory. Financial markets had been fairly stable for some years and a reasonable degree of market efficiency was reached. Nevertheless, investors also exhibited some degree of irrational behavior, which eventually led to the housing bubble and markets exhibiting higher degrees of inefficiency. Since mortgages had been transformed into investment vehicles sold across the globe, the housing crisis quickly evolved into a global financial crisis. Investors had to learn from their mistakes and needed to adapt to the new market conditions. Those investors that did not learn quickly enough and/or did not adapt to the new market situation lost so much money that they were driven out of the market. The new market ecology is now comprised of "old" investors that learned and adapted rapidly and the "new" investors that entered the market after the housing crash. Given these new market participants and conditions, a new evolution towards efficient markets was started.

Lo's framework is in line with the philosophy behind the alternative test methodologies because efficiency is approached as a time-variant characteristic of stock markets. This combination of both an alternative theoretical framework and matching test methodologies can pave the way for academic consensus on the efficiency of financial markets. Early patterns observed by Lo (2004, 2005) seem to provide support for the idea of adaptive and time-varying market efficiency. However, the observed efficiency pattern seems to be opposite to the one expected under the AMH framework. This observation is also in line with what was found by Kim et. al (2011). Part of the explanation might be the imperfect relation between predictability and efficiency, causing predictability proxies of efficiency to be biased.

Despite this critique, some recent research was published in defense of the use of return predictability to examine stock market efficiency. Campbell and Yogo (2006) improve the predictability testing battery by developing a pretest to overcome invalid statistical inference. Cochrane (2008) takes up the defense of return predictability. Greenwood and Shleifer (2013) confirm that investors' expectations are not consistent with results from efficiency proxies. Alternatively, they introduce an innovative behavioral model that explains changing market prices using interactions between different groups of investors. Further empirical research using state-of-the-art time-varying weak form market efficiency tests is needed to shed more light on the validity of the AMH theory.

1.6 Conclusion

The debate on efficient markets has come a long way. In fact, many of the most renowned 19^{th} and 20^{th} century economists have contributed to it to some extent. In our paper, we organize an overview of the theoretical and the empirical debate on market efficiency, reflecting the views of both the advocates of the EMH and behavioral finance. Furthermore, both traditional and alternative test methodologies for weak form market efficiency have been reviewed.

The application of most of these test methodologies has led to conflicting results; a definitive view on market efficiency remains to be found. Lo (2004, 2005) attempted to reconcile both views by means of his adaptive markets hypothesis. Empirically testing the validity of his framework, however, some inconsistencies in the predicted pattern of efficiency are found.

From these findings, we suggest that further research is conducted on two levels. First, more research is needed to further develop the time-varying test methodologies, as these have the potential to circumvent the all-or-nothing discussion on efficiency. Secondly, these empirical developments need implementation, as the AMH should be tested using state-of-the-art methodologies to further look into the documented discrepancies in the pattern of efficiency.

Chapter 2

Towards a new framework on efficient markets

Abstract

Academic research on the efficiency of financial markets goes back several decades. Empirical evidence is mixed and academia is torn between two opposing convictions: the efficient market hypothesis (EMH) vs. behavioral finance. The recent Nobel Prize awarded to scholars from both sides of the debate confirms the stalemate. We apply multiple state-of-the-art efficiency tests in rolling windows of one year to leading global stock market indices to test the adaptive markets hypothesis (AMH), a proposed reconciling framework. We find the idea of dynamic and time-variant efficiency to be valid. Also the theoretical pattern of efficiency predicted by the AMH is in line with our results. Furthermore, we find that the effect of the most recent financial crisis on weak form market efficiency is most prominent on the U.S. stock market. The European and Japanese markets appear more consistently efficient over the course of the last 15 years.

2.1 Introduction

For more than 50 years, researchers have been debating about the informational efficiency of stock markets. Even during the 2013 lectures for the Nobel Prize in Economic Sciences, Fama and Shiller, representing respectively the efficient market hypothesis (EMH) and behavioral finance end of the spectrum, presented opposing evidence on the efficiency of stock markets. Given the renewed appraisal of research on stock market efficiency and the remarkable situation in which two opposing views seem to be irreconcilable, we expand upon earlier empirical work testing an alternative framework on efficiency using a series of tests across the global developed stock markets.

Fama (1970, p. 383) defines an efficient market as “a market in which prices always fully reflect available information” and makes a distinction between different types of efficient markets based on three concretions of the concept “available information” i.e. weak form efficient markets (historical price information), semi-strong form efficient markets (all publicly available information), and strong form efficient markets (all information, both public and private). Following the establishment of the efficient market hypothesis (EMH) by Fama (1970), two schools of thought started to form. On the one hand, proponents of the EMH argue that financial markets are perfectly capable of aggregating information of all investors, which in turn leads to efficient markets. If the price of a stock would appear to be too high given past price information, rational investors would bid the price down to make a profit and vice versa. On the other hand, some researchers started looking into the psychology of investors. In close collaboration with psychologists, the field of behavioral finance was established. Proponents of behavioral finance believe that investors are not always fully rational and therefore are not able to force the stock market to be efficient at all times (e.g. Shefrin, 2000). The debate between these two schools of thought is still going on. The U.S. housing bubble, which eventually triggered the current sovereign debt crisis, sparked newfound interest in this matter. Behaviorists even argue that the EMH can be considered one of the causes of the current financial downturn as policy makers, banks and investors were blindly following the bullish market, while irrational exuberance was building up underneath (Shiller, 2000). More recently, the shared 2013 Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel between scholars with opposing views on efficiency indicated that the debate is far from settled.

We believe that the lack of an alternative theoretical framework is one of the main reasons why the debate on market efficiency yet remains to be settled. One could argue that being critical of an existing theoretical framework is somewhat straightforward. Indeed, a theory is supposed to be imperfect as it is only a framework to describe reality. However, coming up with a new and improved theory is far less evident. Thus far, advocates of behavioral finance have failed in coming up with such a new theory that could replace the EMH, although several biases of behavioral nature have been documented in the academic literature. Following the construct of a time-varying degree of market efficiency (Campbell, Lo, & MacKinlay, 1997) and trying to reconcile theories of the EMH and behavioral finance, Lo (2004, 2005) came up with the adaptive markets hypothesis (AMH). Starting from the concepts of bounded rationality and satisficing[1], and the notion of biological evolution, he argues that many of the biases found in behavioral finance follow a certain evolutionary path, in which individuals try to learn and adapt to new market conditions. This learning and adaptation process is driven by competition among investors, and natural selection determines the new market ecology, with some investors being driven out of the market and some investors remaining in the market. The process of natural selection and competition also shapes the evolutionary dynamics that occur in the market, which are mirrored in the degree of

[1] Humans do not have the information, nor the methodology to always optimize in a rational way. Consequently, they use some rules of thumb or heuristics to find satisfactory results that are not necessarily rational (Simon, 1955).

efficiency of the market. As long as there is no shock that causes market ecology to change, stock markets are fairly efficient. Once a certain event triggers the process of competition and natural selection, markets become temporarily less efficient. When the new market ecology is formed, efficiency of financial markets returns to pre-shock levels. Several elements of Lo's theory can also be recognized in the development of the 2008 financial crisis.

Although potentially inadequate, we cannot help but notice that, to this day, the EMH is still standing. Every introductory course to financial markets still covers the EMH, while alternative theories like the AMH remain underexposed. The AMH was also not discussed by Fama and Shiller in their Prize Lecture on December 8$^{\text{th}}$, 2013 in Stockholm, Sweden. One of the reasons for this might be the limited extent to which the AMH has been tested empirically. Computing rolling first-order autocorrelations of monthly returns as a measure of market efficiency, Lo (2004, 2005) finds a cyclical pattern through time, which confirms the idea of underlying dynamics to the degree of market efficiency. However, Lo's estimated rolling autocorrelation measures are not in line with the idea of markets being relatively efficient for a long time, until a market crash causes a short period of relatively lower efficiency. Rather, his empirical evidence points towards the reverse. In later years, researchers examined the AMH by means of trading strategies. Investigating the profitability of moving average strategies on the Asia-Pacific financial markets, Todea, Ulici, and Silaghi (2009) confirm the cyclical efficiency pattern of the AMH. Neely, Weller, and Ulrich (2009) study excess returns earned by various technical trading rules on foreign exchange markets. They find these returns to decline over time, but at a slower pace than expected under the EMH because of behavioral and institutional factors. These findings are consistent with the AMH view of markets being dynamic systems subject to underlying evolutionary processes. Tests of the AMH have also been conducted on markets other than the stock market. For example, Zhou and Lee (2013) confirm the underlying propositions of the AMH for real estate investment trusts, using two different weak form efficiency tests. Finding a higher degree of stock market predictability in times of economic and political crises, Kim, Shamsuddin, and Lim (2011) confirm Lo's idea of time-varying market efficiency being driven by changing market conditions. During market bubbles and crashes, virtually no return predictability is found. This, however, is at odds with Lo's AMH, which states that higher degrees of predictability and thus lower degrees of efficiency ought to be found in times of market mania. Urquhart and Hudson (2013) find the AMH to provide a better description of stock market efficiency than the EMH, using a number of standard tests for efficiency in five-year rolling windows on global stock markets. Finally, Lim, Luo and Kim (2013) test return predictability on U.S. stock markets using two correlation-based tests and find the time-varying nature of return predictability to be consistent with the AMH.

The first evidence from empirical studies shows that there is value to the idea of adaptive markets, but some discrepancies were found as well. However, these studies have been rather limited in terms of applied methodologies and/or the geographical variety of financial markets considered. To complement the literature, we apply six state-of-the-art rolling efficiency tests on three leading stock market indices from around the developed world to gain more insights into the validity of Lo's theory. From the results, we are able to validate both the concept of time-varying efficiency and the anticipated efficiency pattern. However, the evidence is most compelling for linear tests of U.S. stock market efficiency, and more limited with respect to the last 15 years for the European and Japanese markets.

2.2 Data and methodology

2.2.1 Data

When implementing rolling efficiency tests, some decisions need to be made on the structure of the underlying data. Specifically, different data frequencies and window lengths could be considered. To avoid sensitivity of test results to these two data preprocessing decisions, we first performed a robustness analysis from which we learn that daily data and rolling windows of length 1 year are most robust (Verheyden, 2013), in contrast to longer windows.

We collect data through Thomson Reuters Datastream on the daily prices of three leading indices across the global developed markets. The Standard & Poor's 500 stock market index represents 500 of the largest companies in the United States and is known as the leading indicator for the broad U.S. stock market. The Euro Stoxx 50 is the leading blue-chip index representing the Eurozone. The Nikkei 225 is the prominent Japanese stock market index that represents the developed Asian economies. Daily prices are transformed to continuously compounded daily returns to implement the different methods. The first observation included for every index is the base date in Datastream. The last included observation is the return on December 20$^{\text{th}}$, 2013. The evolution of the daily returns on the different indices is plotted in Figure 2.1. Summary statistics of the return data are presented in Table 2.1.

2.2.2 Methodology

To revisit the empirical results on the market efficiency debate, we implement six state-of-the-art empirical tests that are being used in the literature to test for return predictability and weak form market efficiency. These

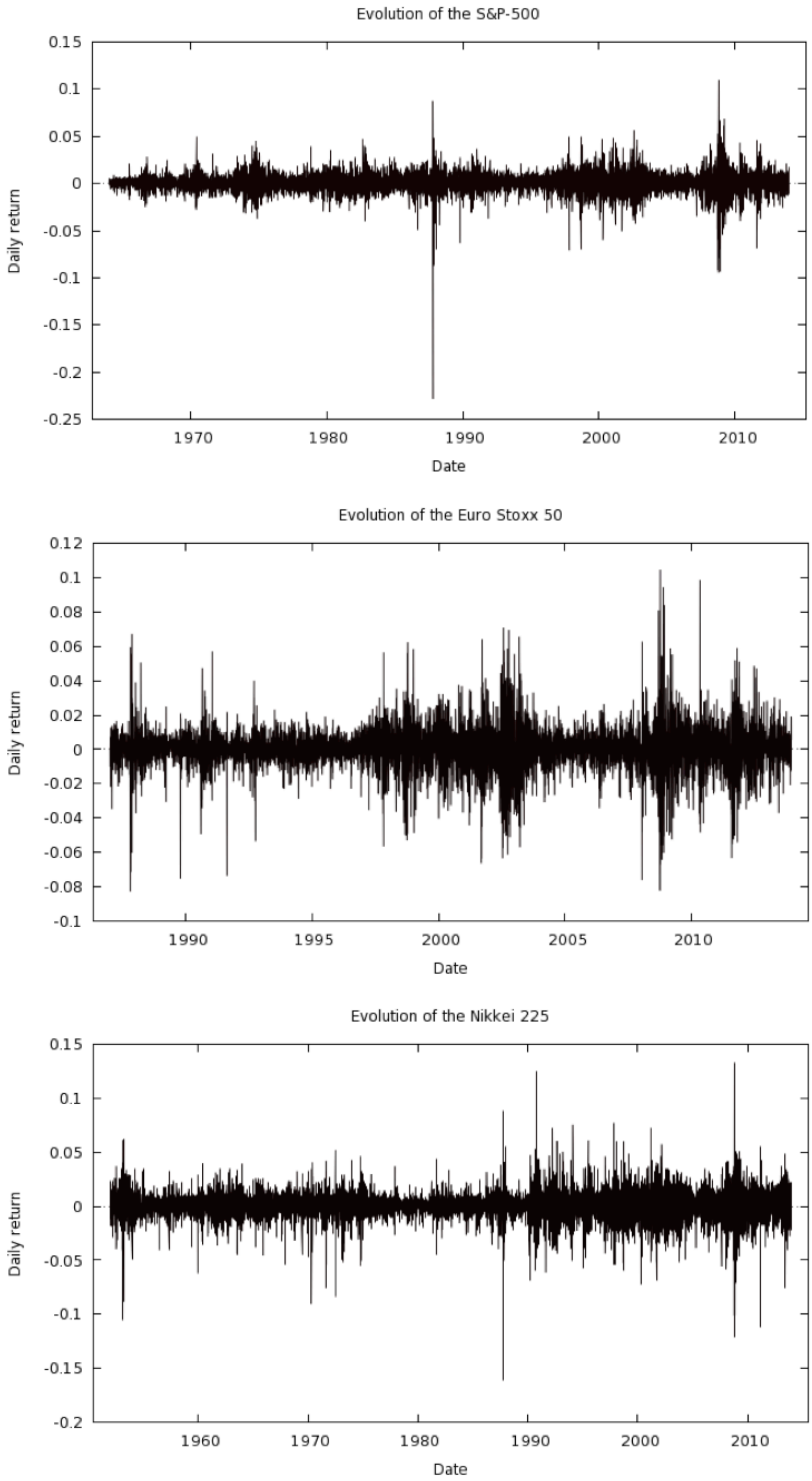


Figure 2.1: *Evolution of daily returns*

Summary statistics			
	S&P-500	*Euro Stoxx 50*	*Nikkei 225*
Begin	1/01/1964	1/01/1987	1/01/1951
End	20/12/2013	20/12/2013	20/12/2013
Mean	0.00024	0.00017	0.00028
Median	0.00013	0.00038	0.00004
Minimum	-0.22833	-0.08262	-0.16135
Maximum	0.10957	0.10438	0.13235
Standard deviation	0.01017	0.01327	0.01182
Coefficient of variation	41.58900	76.60000	41.92600
Skewness	-1.03580	-0.14949	-0.49094
Excess kurtosis	28.08600	6.02940	10.96200
5% percentile	-0.01514	-0.02095	-0.01815
95% percentile	0.01515	0.01947	0.01787
Interquartile range	0.00917	0.01221	0.01067

Summary statistics of daily returns of the S&P-500, the Euro Stoxx 50 and the Nikkei 225.

Table 2.1: *Summary statistics*

tests are applied on leading stock market indices representing the United States, the Eurozone and Japan, to verify the validity of the AMH across the developed world. Note that we implement all tests in a rolling version, using window lengths of 1 year. The choice for rolling window tests to investigate the AMH follows naturally from the concept of time-varying efficiency. In this section, we briefly review the applied methods, and refer to the relevant literature for a more in-depth discussion.

Wild bootstrap automatic variance ratio test

An established method to examine weak form market efficiency is through variance ratio (VR) tests. This type of tests goes back to the work of Lo and MacKinlay (1988), who established a first VR that proved to be very popular to test for an uncorrelated increment and that can also be used to determine whether or not a stock market is weak form efficient over a certain period of time. The main assumption behind the test is that if stock returns follow an uncorrelated increment (random walk 3), the variance of the stock returns over a certain time interval s is the same as k times the variance of the stock returns over an interval s/k^2. Statistically, this simple yet elegant relationship can be tested by calculating if the ratio of the variance of $r_t + r_{t-1} + \cdots + r_{t-k+1}$ over k times the variance of r_t significantly differs from unity. The null hypothesis of this test states that the time-series is following an uncorrelated increment model. Whenever the ratio statistically differs from unity, the null hypothesis can be rejected and we arrive at the alternative hypothesis stating that the time series is not following a random walk (version 3).

Assuming that r_t is the return on a certain stock at time t ($t = 1, \cdots, T$), the Lo and MacKinlay VR test statistic is calculated as follows:

$$VR(k) \equiv \frac{Var(r_t(k))}{kVar(r_t)} = 1 + 2\sum_{j=1}^{k-1}(1 - \frac{j}{k})\rho_j \tag{2.1}$$

With $r_t(k) \equiv r_t + r_{t-1} + \cdots + r_{t-k+1}$ and ρ_j being the j^{th} order autocorrelation coefficient of r_t.

Over the years, the Lo-MacKinlay VR test became critiqued and has been improved numerous times. One improvement is the use of a data dependent procedure to avoid arbitrary determination of the vector of holding periods k (Choi, 1999):

$$AVR(\hat{k}) = \sqrt{\frac{T}{k}}\frac{[VR(\hat{k}) - 1]}{\sqrt{2}} \xrightarrow{d} N(0, 1) \tag{2.2}$$

[2] For example: under the uncorrelated increment model, the variance of stock returns over a time interval of 10 weeks will be statistically indifferent from 10 times the variance of stock returns over a time interval of 1 week.

A further improvement is the use of wild bootstrapping to take into account possible conditional heteroscedasticity in returns, particularly in small samples (Kim, 2009; Charles, Darné, & Kim, 2011). A wild bootstrap version of the automatic variance ratio test (AVR), which results in a test statistic and the associated rolling confidence intervals, is considered state-of-the-art. Note that we use 500 bootstrap iterations to perform the test.

Power transformed joint variance ratio test

Next to the wild bootstrap automatic variance ratio test (Kim, 2009; Charles et al., 2011), two other variance ratio tests prove to be popular for empirical examination of the weak form market efficiency hypothesis. A first alternative is the power transformed joint variance ratio (Chen & Deo, 2006). The main drawback of the standard Lo-MacKinlay (1988) VR test is the right skewed null distribution of the test statistic in finite samples, as opposed to the theoretically suggested normal distribution. The proposed power transformation by Chen and Deo is able to solve this right skewness problem, and also proves to be robust for possible conditional heteroscedasticity in the return series. A vector of holding periods is required as an input to this method. We choose holding periods of 2, 5 and 10 days, as advocated by Deo and Richardson (2003). The obtained test statistic follows a χ^2 distribution under the null hypothesis of a random walk. As for the remaining tests, we do not present the Chen-Deo test statistic in more detail, but instead refer to the original paper and the excellent review paper on variance ratio tests of the random walk by Charles and Darné (2009).

Chow-Denning multiple variance ratio test

Another state-of-the-art VR alternative for the Lo-MacKinlay (1988) test is the multiple VR test (Chow & Denning, 1993), which has often been applied in the empirical literature as well. The traditional Lo-MacKinlay VR test is an individual test, where the test statistic is calculated for individual values of the holding period k. To examine weak form market efficiency, such a test needs to be conducted separately multiple times for different values of k, which leads to an over rejection of the null hypothesis, and thus a larger type I error. This is referred to as a multiple testing phenomenon. Chow and Denning stressed this problem and proposed a heteroscedasticity-robust multiple variance ratio test, following earlier work of Hochberg (1974). A vector of individual Lo-MacKinlay test statistics for pre-defined holding periods is constructed, while controlling for overall test size by means of studentized maximum modulus critical values. Again, we adopt a holding period vector of 2, 5 and 10 days, following Deo and Richardson (2003). In this multiple variance ratio test, the nullity of a random walk is rejected as soon as any of the m test statistics included in the vector significantly deviates from unity.

Belaire-Franch and Contreras test

As mentioned earlier, one of the issues with the Lo-MacKinlay (1988) VR test is the right skewed sampling distribution. Amongst others, Chen and Deo (2006), Kim (2009) and Charles et al. (2011) provide a parametric solution to this problem. Alternatively, Wright (2000) suggests a non-parametric alternative using rank and sign test statistics, which have an exact sampling distribution and which appear to be more powerful in the presence of serial correlation. The sign-based test is even exact under conditional heteroscedasticity. Belaire-Franch and Contreras (2004) further improved the Wright test by transforming it to a multiple test in the same fashion as Chow and Denning (2003) did with the Lo-MacKinlay VR test. Holding periods of 2, 5 and 10 days are used, which is in line with suggestions from earlier research (Deo & Richardson, 2003). We focus on the critical values for the joint sign test, which are obtained through simulation based on sample size and a vector of holding periods.

Automatic portmanteau test

Another popular way to test for weak form market efficiency is through joint tests of serial correlation in the return series. A standard way of doing this is through the Ljung-Box (1978) test statistic, with the null hypothesis that the first k serial correlations are statistically indistinguishable from zero:

$$Q_p = T \sum_{k=1}^{p} \hat{\rho}_k^2 \tag{2.3}$$

With T the number of return observations and p the largest order of serial correlation included in the test statistic. As a general rule of thumb, p is often taken as $\sqrt{T}$. Given that the test statistic is a sum of squared normals, critical values are taken from a χ^2 distribution.

The above test statistic is still flawed in two ways: it assumes independence of returns and the determination of p is rather arbitrary. Escanciano and Lobato (2009) propose a robust automatic portmanteau test that addresses both flaws. The test statistic remains virtually the same, but the optimal lag p is determined using both the Akaike information criterion (AIC) and the Bayesian information criterion (BIC).

Generalized spectral test

Next to the previous tests based on linear serial correlation, it is of interest to also test for non-linear serial dependence in the data structure as a measure of weak form market efficiency. Escanciano and Velasco (2006) propose a generalized spectral test for the martingale difference hypothesis (MDH). Under the MDH, the best prediction of future values of a time series is simply the unconditional expectation, which points to weak form market efficiency. Also, the normalized spectral density function of returns is equal to one across all frequencies, which is a central property in the calculation of the generalized spectral test. Given that this type of test statistic does not follow a standard asymptotic distribution, a wild bootstrapping procedure, which also takes into account possible conditional heteroscedasticity, should be implemented. The null hypothesis of the test states that no return predictability, neither linear nor non-linear, is present in the return series, which hints at weak form efficiency. If the obtained p-value is 0.05 or higher, the null hypothesis cannot be rejected with a confidence of 95%. Note that we perform 300 bootstrap iterations to execute the test.

2.3 Results

To perform the different return predictability tests, we make use of the *vrtest* library in R. The tests are applied to daily data in rolling windows of 260 observations (1 year). Applying the tests to rolling windows of the dataset and plotting the results from each individual window yields a time-varying measure for predictability, which could in turn be interpreted as a time-variant degree of weak form market efficiency. We collect the test statistics (p-value for the generalized spectral test) and the associated 95% critical values and approximate the date by the middle observation between the beginning and the end of every window.

An interesting way of summarizing the data is by calculating an efficiency ratio, which is the fraction of observations for which the null hypothesis of no return predictability cannot be rejected over the total number of observations. In turn, this ratio can be interpreted as a proxy for the degree of weak form market efficiency. The efficiency ratios across the different tests for the three indices are displayed in Table 2.2.

On average, the S&P-500 appears to be weak form efficient around 75% of the time from January of 1964 to December of 2013. Overall, the efficiency ratios across the different tests are in the same order of magnitude, which adds robustness to the results. However, we do note that the generalized spectral test is pointing to relatively higher efficiency for the S&P-500. Overall weak form efficiency of the Euro Stoxx 50 seems higher, although we need to take into account the difference in sample length. The same goes for the Nikkei 225.

To accommodate the comparison of efficiency ratios across different indices, we also compute the corresponding efficiency ratios (Table 2.3) for the sample period from January 1$^{\text{st}}$, 1987 until December 20$^{\text{th}}$, 2013. The overall efficiency levels are now relatively higher for the S&P-500 and the Nikkei 225, which could suggest that weak form market efficiency has grown over the years, as stock market technology has improved. On average, overall efficiency levels appear to be lowest in the U.S. (90.1%); Europe has an overall average weak form market efficiency level of 93%; Japan appears to be most efficient with an overall average of 96.4%.

The complete rolling weak form efficiency graphs for the S&P-500, Euro Stoxx 50 and Nikkei 225 across the different methodologies are plotted resp. in Figure 2.2, 2.3 and 2.4. A first observation from the different plots is the varying degree of efficiency through time. For both the S&P-500 and the Nikkei 225 we notice significantly higher degrees of return predictability prior to the 1980s. Hence, the degree of weak form market efficiency was lower. From the 1980s onward, all three indices can generally be considered weak form efficient, with the exception of some shorter-term

Efficiency ratios	
S&P-500 (01/01/1964 - 20/12/2013)	
Wild bootstrap automatic variance ratio test	68.3%
Automatic Portmanteau test	65.4%
Power transformed joint variance ratio test	73.5%
Chow-Denning multiple variance ratio test	76.4%
Belaire-Franch and Contreras joint sign test	71.2%
Generalized spectral test	95.5%
Euro Stoxx 50 (01/01/1987 - 20/12/2013)	
Wild bootstrap automatic variance ratio test	95.3%
Automatic Portmanteau test	87.9%
Power transformed joint variance ratio test	96.4%
Chow-Denning multiple variance ratio test	97.8%
Belaire-Franch and Contreras joint sign test	84.2%
Generalized spectral test	96.4%
Nikkei 225 (01/01/1951 - 20/12/2013)	
Wild bootstrap automatic variance ratio test	81.4%
Automatic Portmanteau test	86.4%
Power transformed joint variance ratio test	94.9%
Chow-Denning multiple variance ratio test	91.0%
Belaire-Franch and Contreras joint sign test	78.4%
Generalized spectral test	95.0%

Efficiency ratios from the different empirical tests for the daily returns of the S&P-500, Euro Stoxx 50 and Nikkei 225.

Table 2.2: *Efficiency ratios*

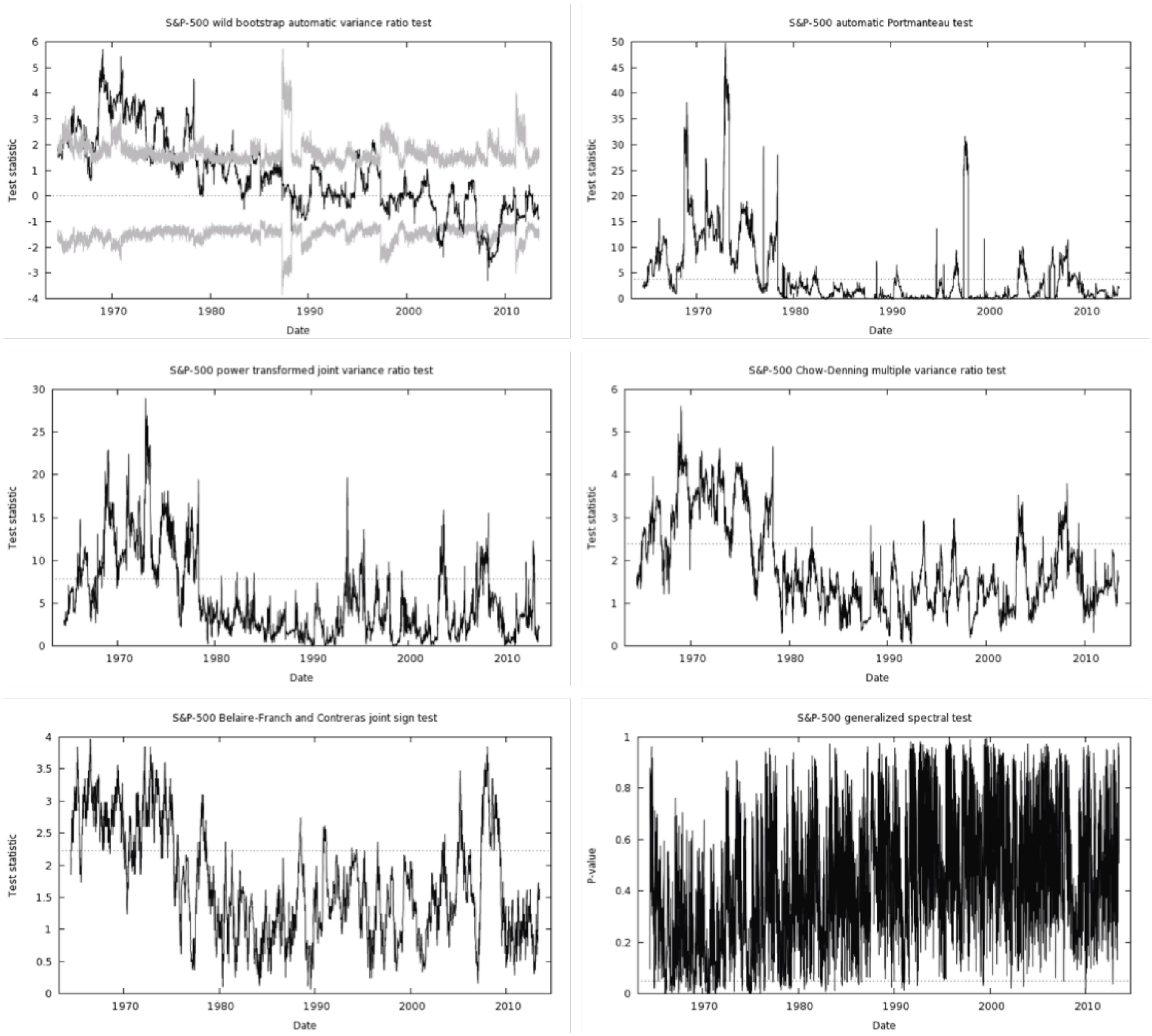
Corresponding efficiency ratios	
S&P-500 (01/01/1987 - 20/12/2013)	
Wild bootstrap automatic variance ratio test	85.3%
Automatic Portmanteau test	82.9%
Power transformed joint variance ratio test	90.7%
Chow-Denning multiple variance ratio test	91.0%
Belaire-Franch and Contreras joint sign test	91.0%
Generalized spectral test	99.5%
Euro Stoxx 50 (01/01/1987 - 20/12/2013)	
Wild bootstrap automatic variance ratio test	95.3%
Automatic Portmanteau test	87.9%
Power transformed joint variance ratio test	96.4%
Chow-Denning multiple variance ratio test	97.8%
Belaire-Franch and Contreras joint sign test	84.2%
Generalized spectral test	96.4%
Nikkei 225 (01/01/1987 - 20/12/2013)	
Wild bootstrap automatic variance ratio test	95.4%
Automatic Portmanteau test	96.2%
Power transformed joint variance ratio test	99.5%
Chow-Denning multiple variance ratio test	99.0%
Belaire-Franch and Contreras joint sign test	90.3%
Generalized spectral test	98.1%

Corresponding efficiency ratios from the different empirical tests for the daily returns of the S&P-500, Euro Stoxx 50 and Nikkei 225 for the period between January 1^{st}, 1987 and December 20^{th}, 2013

Table 2.3: *Corresponding efficiency ratios*

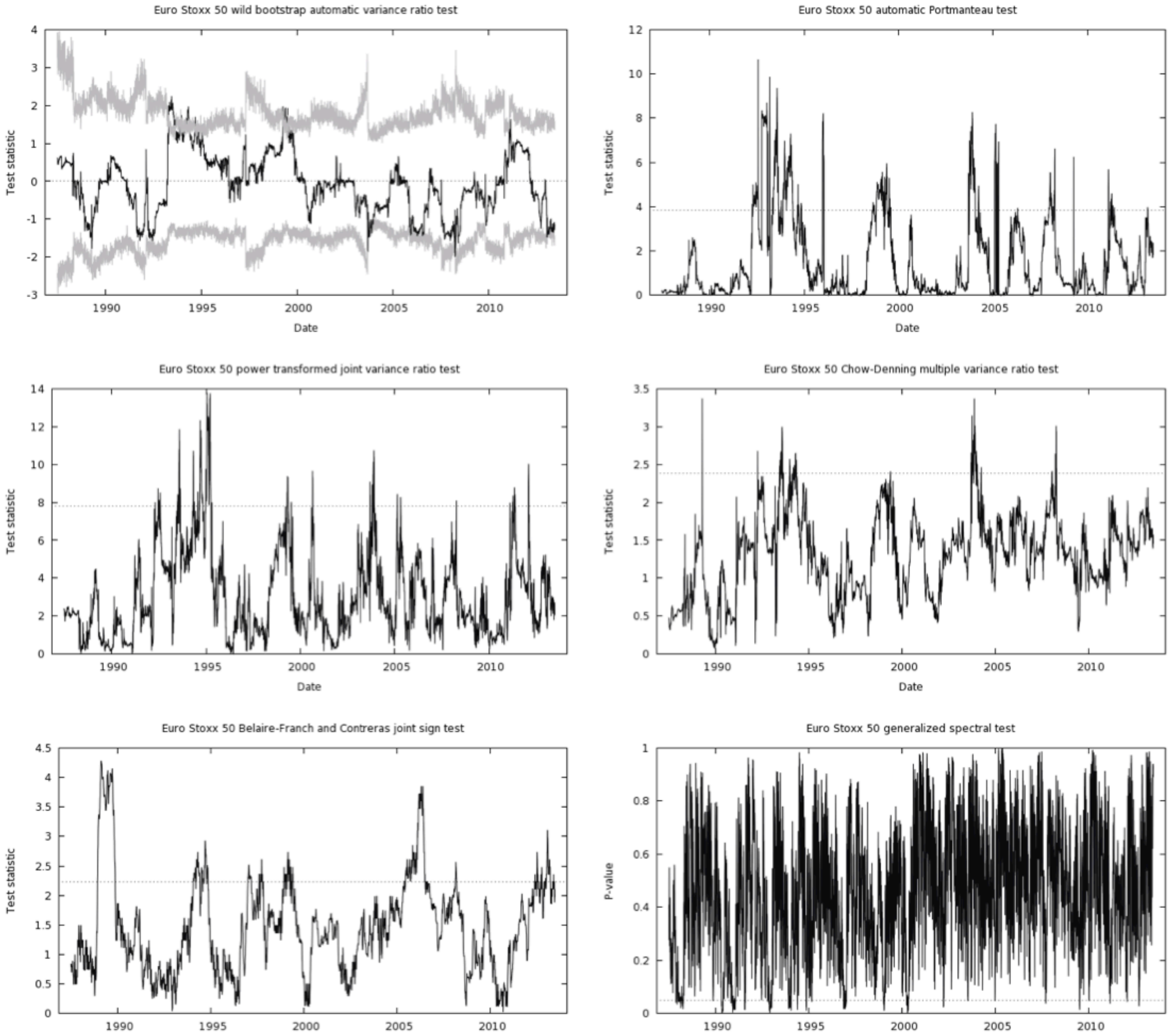
periods with higher return predictability. Table 2.4, 2.5 and 2.6 present a better overview of those periods where the null hypothesis of no return predictability is rejected for the different tests with 95% confidence. Note that we only include those periods where the null hypothesis is rejected for two or more consecutive months. This enables us to distinguish the efficient market equilibrium from times with a significantly lower degree of weak form market efficiency, rather than just very short-term deviations.

From this summary, we confirm the idea that the overall market was generally less weak form efficient prior to the 1980s. Post 1980s, the S&P-500 is rather weak form efficient, with the exception of lower degrees of weak form market efficiency in 2003 and in the buildup to the most recent financial crisis. For the Nikkei 225, we confirm the same trend of lower degrees of market efficiency prior to the 1980s, even though this pattern is less convincing given relatively shorter periods of lower market efficiency during the 1960s and 1970s in comparison to the S&P-500. No data prior to 1987 is included for the Euro Stoxx 50, which leads to a smaller amount of periods with significantly lower degrees of market efficiency. Apparently, the European market faced some deviations from market efficiency during the 1990s and to a lesser extent during the period leading to the most recent financial crisis that started in 2008. The Japanese market reflects similar deviations over the course of the 1990s, but market efficiency seems unaffected prior to and during the 2008 financial turmoil. This result is quite striking, and shows that the effects of the most recent financial crisis were mainly reflected in the U.S. stock market efficiency, and to a lesser extent in Europe and Japan. Note that consecutive rejections of weak form efficiency from the generalized spectral test are more dispersed in time and never exceed two months in length.



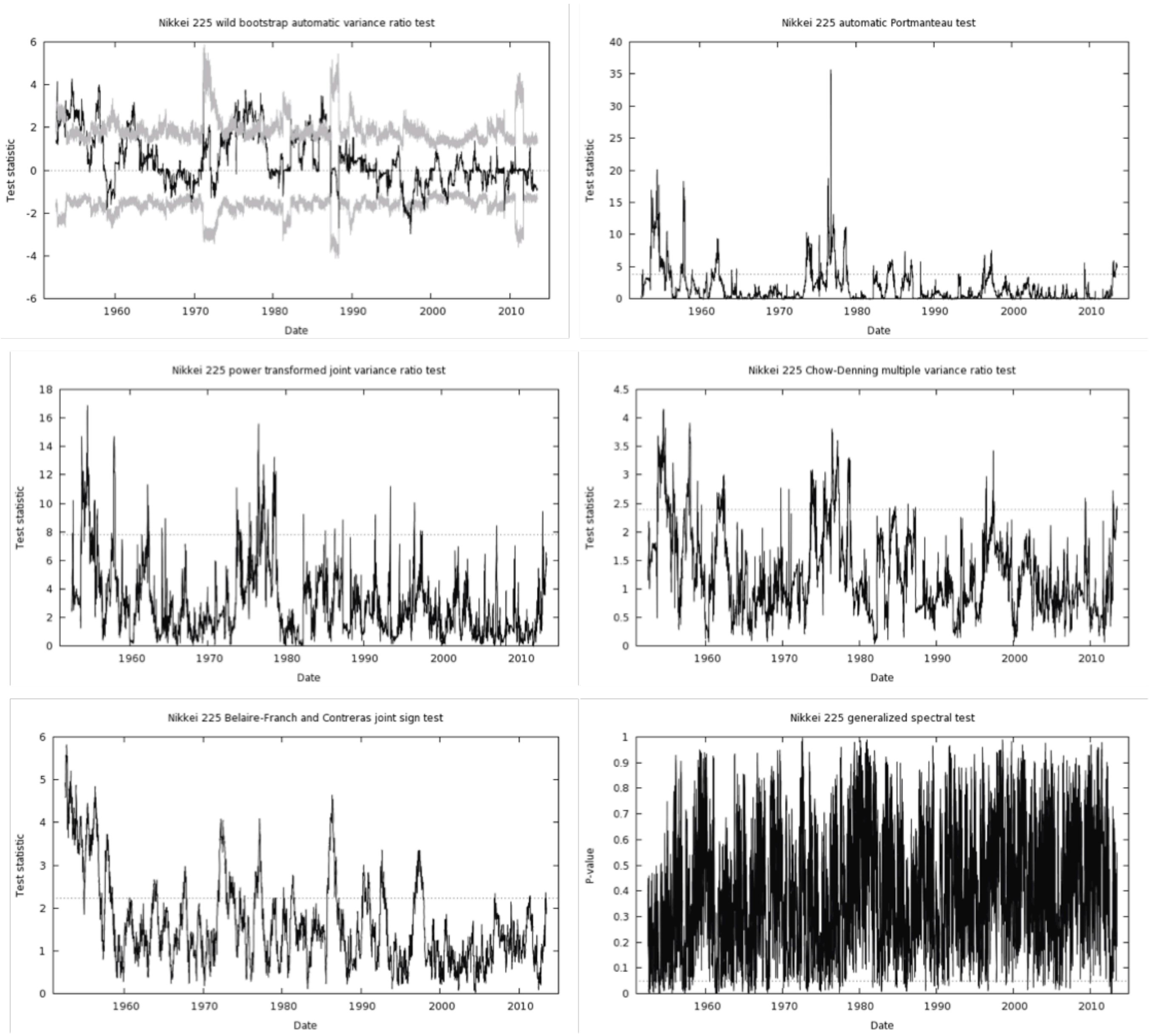
Six methods were implemented on daily return data of the S&P-500 (01/01/1964 – 20/12/2013): wild bootstrap automatic variance ratio test, automatic Portmanteau test, power transformed joint variance ratio test, Chow-Denning multiple variance ratio test, the Belaire-Franch and Contreras joint sign test and the generalized spectral test. For every method, both the test statistic and the 95% critical value(s) are plotted. For the wild bootstrap automatic variance ratio test, the market is weak form efficient when the calculated test statistic falls within the 95% bootstrapped confidence interval. With the other tests, the market is considered efficient when the estimated test statistic is smaller than or equal to the associated critical value.

Figure 2.2: *Rolling return autocorrelation of the S&P-500*



Six methods were implemented on daily return data of the Euro Stoxx 50 (01/01/1987 – 20/12/2013): wild bootstrap automatic variance ratio test, automatic Portmanteau test, power transformed joint variance ratio test, Chow-Denning multiple variance ratio test, the Belaire-Franch and Contreras joint sign test and the generalized spectral test. For every method, both the test statistic and the 95% critical value(s) are plotted. For the wild bootstrap automatic variance ratio test, the market is weak form efficient when the calculated test statistic falls within the 95% bootstrapped confidence interval. With the other tests, the market is considered efficient when the estimated test statistic is smaller than or equal to the associated critical value.

Figure 2.3: *Rolling return autocorrelation of the Euro Stoxx 50*



Six methods were implemented on daily return data of the Nikkei 225 (01/01/1951 – 20/12/2013): wild bootstrap automatic variance ratio test, automatic Portmanteau test, power transformed joint variance ratio test, Chow-Denning multiple variance ratio test, the Belaire-Franch and Contreras joint sign test and the generalized spectral test. For every method, both the test statistic and the 95% critical value(s) are plotted. For the wild bootstrap automatic variance ratio test, the market is weak form efficient when the calculated test statistic falls within the 95% bootstrapped confidence interval. With the other tests, the market is considered efficient when the estimated test statistic is smaller than or equal to the associated critical value.

Figure 2.4: *Rolling return autocorrelation of the Nikkei 225*

"Inefficiency periods" S&P-500					
Wild bootstrap automatic VR test			*Power transformed joint VR test*		
1/11/1965	-	26/02/1966	25/12/1965	-	26/02/1966
6/04/1966	-	20/11/1966	28/04/1966	-	4/12/1966
19/05/1968	-	16/06/1973	5/12/1967	-	30/04/1968
31/03/1974	-	6/10/1975	21/05/1968	-	26/11/1969
13/04/1977	-	14/05/1978	18/02/1970	-	26/09/1973
17/06/1980	-	27/09/1980	29/09/1973	-	21/11/1973
5/12/1994	-	3/04/1995	19/03/1974	-	15/02/1976
16/06/1996	-	6/11/1996	2/11/1976	-	30/10/1977
10/04/2003	-	2/11/2003	28/11/1977	-	23/01/1978
24/04/2007	-	9/09/2007	25/01/1978	-	6/05/1978
20/09/2007	-	9/04/2008	25/12/1994	-	28/03/1995
16/04/2008	-	27/03/2009	10/04/2003	-	18/09/2003
Automatic Portmanteau test			11/06/2007	-	1/09/2007
31/12/1964	-	16/02/1967	21/09/2007	-	8/04/2008
5/12/1967	-	8/04/1976	*Belaire-Franch/Contreras joint sign test*		
13/04/1977	-	4/05/1978	14/07/1964	-	29/05/1965
21/07/1990	-	6/10/1990	16/08/1965	-	26/07/1968
20/06/1996	-	22/12/1996	1/08/1968	-	21/01/1970
30/06/1997	-	16/12/1997	3/08/1970	-	24/10/1970
1/02/2003	-	5/11/2003	19/04/1971	-	27/09/1971
7/06/2006	-	12/11/2006	17/11/1971	-	30/11/1973
4/05/2007	-	8/04/2008	3/02/1974	-	10/08/1974
22/07/2008	-	15/09/2008	14/08/1974	-	25/03/1975
21/09/2008	-	7/03/2009	4/01/1978	-	8/07/1978
Chow-Denning multiple VR test			24/05/1988	-	2/08/1988
5/02/1965	-	29/05/1965	26/12/1990	-	8/03/1991
13/09/1965	-	22/12/1965	21/01/2005	-	27/06/2005
25/12/1965	-	26/02/1966	17/05/2007	-	6/05/2009
3/03/1966	-	4/12/1966	*Generalized spectral test*		
5/12/1967	-	24/11/1969	/	-	/
27/11/1969	-	10/11/1973			
17/03/1974	-	14/02/1976			
9/05/1977	-	8/05/1978			
10/04/2003	-	21/09/2003			
16/05/2007	-	2/09/2007			
12/09/2007	-	8/04/2008			

A prolonged period is defined as a period in which the null hypothesis of no return predictability is rejected for two or more consecutive months.

Table 2.4: *Summary of prolonged periods of lower degrees of weak form market efficiency for the different tests of the daily S&P-500 returns (01/01/1964 – 20/12/2013)*

"Inefficiency periods" Euro Stoxx 50					
Wild bootstrap automatic VR test			*Automatic Portmanteau test*		
20/04/1993	-	8/08/1993	22/04/1992	-	4/08/1992
Power transformed joint VR test			7/10/1992	-	22/01/1993
31/08/1994	-	30/10/1994	23/05/1993	-	28/08/1993
24/12/1994	-	8/04/1995	23/11/1993	-	27/01/1994
Belaire-Franch/Contreras joint sign test			29/01/1994	-	30/05/1994
5/12/1988	-	16/11/1989	28/10/1998	-	11/03/1999
14/03/1994	-	29/05/1994	30/09/2003	-	22/01/2004
25/12/1996	-	2/03/1997	15/12/2007	-	6/02/2008
20/06/2005	-	8/07/2006	*Chow-Denning multiple VR test*		
4/01/2013	-	26/03/2013	/	-	/
Generalized spectral test					
/	-	/			

A prolonged period is defined as a period in which the null hypothesis of no return predictability is rejected for two or more consecutive months.

Table 2.5: *Summary of prolonged periods of lower degrees of weak form market efficiency for the different tests of the daily Euro Stoxx 50 returns (01/01/1987 – 20/12/2013).*

"Inefficiency periods" Nikkei 225					
Wild bootstrap automatic VR test			*Automatic Portmanteau test*		
28/09/1953	-	24/06/1955	3/09/1953	-	23/05/1955
9/08/1955	-	3/01/1956	23/09/1955	-	25/01/1956
8/07/1957	-	24/01/1958	11/07/1957	-	13/11/1957
24/06/1961	-	26/09/1961	17/11/1957	-	20/01/1958
31/10/1961	-	26/06/1962	7/12/1961	-	24/06/1962
6/06/1975	-	25/09/1975	4/08/1973	-	21/03/1974
27/04/1976	-	6/06/1977	22/04/1976	-	5/06/1977
1/08/1977	-	8/12/1977	3/06/1978	-	10/10/1978
25/05/1978	-	10/10/1978	28/01/1984	-	29/09/1984
4/06/1984	-	29/09/1984	12/01/1987	-	15/03/1987
24/01/1986	-	30/09/1986	4/04/1997	-	20/06/1997
25/12/1986	-	20/04/1987	*Belaire-Franch/Contreras joint sign test*		
24/11/1996	-	1/02/1997	29/06/1952	-	6/11/1954
4/02/1997	-	20/07/1997	2/12/1954	-	19/11/1956
Power transformed joint VR test			8/07/1957	-	17/03/1958
14/10/1953	-	29/01/1954	30/12/1963	-	29/02/1964
1/02/1954	-	15/10/1954	7/06/1967	-	2/11/1967
24/05/1976	-	5/08/1976	10/01/1972	-	23/11/1972
5/01/1977	-	30/04/1977	8/04/1973	-	21/10/1973
3/06/1978	-	10/10/1978	17/07/1976	-	10/06/1977
Chow-Denning multiple VR test			15/03/1981	-	26/06/1981
16/10/1953	-	8/12/1954	17/09/1985	-	27/11/1986
15/07/1957	-	3/11/1957	31/03/1990	-	8/06/1990
30/06/1961	-	23/09/1961	14/11/1990	-	1/02/1991
16/10/1973	-	27/01/1974	5/06/1992	-	7/12/1992
7/05/1976	-	15/09/1976	30/10/1996	-	20/11/1997
19/10/1976	-	9/05/1977	*Generalized spectral test*		
24/06/1978	-	10/10/1978	/	-	/

A prolonged period is defined as a period in which the null hypothesis of no return predictability is rejected for two or more consecutive months.

Table 2.6: *Summary of prolonged periods of lower degrees of weak form market efficiency for the different tests of the daily Nikkei 225 returns (01/01/1951 – 20/12/2013).*

2.4 Discussion

Our empirical investigation of the AMH confirms the dynamic character of weak form market efficiency, as suggested by Lo (2004, 2005). These results are robust across the different methodologies and developed markets. Rather than markets being efficient all the time, there appears to be an evolution in the degree to which markets are efficient in incorporating past price information, i.e. weak form market efficiency. From this finding we learn that it is not worthwhile to discuss whether financial markets are absolutely efficient or not, as is often the case in the debate between believers of the EMH and advocates of behavioral finance. Instead, market efficiency seems to be situated somewhere between both extreme ends of the spectrum, which opens up the possibility for reconciliation.

Looking into the shifts in weak form efficiency and the periods with significant return predictability also yields some valuable insights. Up until the beginning of the 1980s, financial markets across the globe appear to exhibit higher degrees of return predictability, and hence lower degrees of weak form market efficiency. This finding is also robust across the different methods. Technological advances in trading and computer systems provide a valuable explanation (Gu & Finnerty, 2002). Efficiency ratios calculated from the 1980s onward appear to be significantly higher. However, some prolonged periods where the market efficiency null hypothesis can be rejected are present as well. For the U.S. stock market, the most compelling period with higher-than-normal return predictability is found in the months leading up to the failing of Lehman Brothers and the outburst of the 2008 financial crisis. As a consequence, we can see irrationality building up in the market prior to the crash of the housing bubble, which initiated the global financial crisis. Five years later, the U.S. market seems to have settled in an efficient state again, as return predictability is not significantly different from zero. Remarkably, the same crisis did not affect the efficiency of the European and Japanese stock market to the same extent, as deviations from the efficient equilibrium are more short-lived compared to the U.S. stock market. The Japanese stock market even appears to be consistently weak form efficient over the last 15 years, except for some very short-lived aberrations. These observations are in line with most ideas behind the AMH, although the effect of changing market conditions appears to be less constant across different stock indices. Overall, we find evidence that markets can be relatively efficient in incorporating information for a long time, until a certain event causes a disequilibrium in which investors and markets need to adapt. Once this learning period is over, the market can return to its efficient equilibrium (Lo, 2004, 2005).

Like Lo (2005), Kim et al. (2011), Urquhart and Hudson (2013) and Lim et al. (2013) we confirm the concept of time-varying market efficiency. Together this evidence makes a compelling case to change the tone of the efficient market debate and the direction of future research. The dynamic character of market efficiency has also been confirmed by research implementing trading strategies (Todea et al., 2009; Neely et al., 2009) and research focusing on other types of markets (e.g. Zhou & Lee, 2013). Our results add to this existing literature as well. In contrast to Lo (2005) and Kim et al. (2011), we confirm the intuition on market efficiency in the AMH, as we find that financial turmoil and changing market conditions coincide with periods of lower degrees of market efficiency. Using more sophisticated and state-of-the-art tests, and more robust subsamples of length 1 year, we also confirm most findings from Urquhart and Hudson (2013). Complementary to Lim et al. (2013) we also consider the European and Japanese market and non-linear correlation-based tests. We can discern the effect of the most recent financial crisis between the three leading stock market indices from around the developed world. Our results are robust across a series of state-of-the-art methodologies and across the global developed financial markets. Note however that results from the generalized spectral test point to markets being slightly more efficient, and aberrations being more short-lived and dispersed in time. A possible explanation for this observed behavior can be the non-linear predictability that is accounted for, in contrast to the linear tests of efficiency.

2.5 Conclusion

For many years, scholars have been debating about the efficiency of financial markets. As no consensus can be found, two views on efficiency are prevailing. On the one hand, many still adhere to the EMH and the assumption that markets are able to efficiently incorporate past and public price information. On the other hand, behaviorists argue that investors suffer from psychological anomalies, which introduces irrationality and pushes market prices away from the rational and efficient underlying fundamental value. The stalemate in this area of research was recently illustrated by the Nobel Committee jointly awarding Fama and Shiller, together with Hansen for his work on uncertainty, the 2013 Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel. Most remarkably, both Fama and Shiller openly contradicted each other's beliefs and empirical evidence during their Nobel Prize lecture.

The lack of a widely supported theoretical alternative for the EMH is one of the reasons the debate on efficient markets yet remains to be settled. Lo's AMH (2004, 2005) attempts to address this lack by reconciling between theories of the EMH and behavioral finance, drawing from concepts of evolutionary biology. Neither Fama nor Shiller mentioned this possibly reconciliating framework, which might be due to the inconclusive empirical evidence that has been found so far. However, implementing six state-of-the art rolling tests for return predictability on the three largest developed financial markets (U.S., Europe and Japan), we can validate the central ideas of the AMH. We corroborate the time-varying character of weak form market efficiency and confirm the intuition on the pattern of efficiency predicted in Lo's AMH, in contrast to earlier research. The U.S. stock market appears to be most sensitive to changes in efficiency; the European and Japanese markets behaved more consistently efficient over the last 15 years.

From our findings, we suggest that future research is geared towards the further examination and development of the theory on adaptive markets, following Lo (2004, 2005). The ongoing debate could significantly benefit from such research, as both extreme views can be reconciled. Furthermore, our results illustrate that rolling tests of return predictability can point to inefficiency and irrationality building up in the market, which eventually leads to market crashes. Consequently, these tests might be instrumental in predicting asset bubbles and future turmoil in the financial marketplace.

Chapter 3

Mutual fund performance: A market efficiency perspective

Abstract

Following years of academic debate between proponents and opponents of the efficient market hypothesis (EMH), more and more empirical evidence is now pointing towards an alternative theoretical framework, i.e. the adaptive markets hypothesis (AMH), which predicts time-varying degrees of market efficiency. Following this reconciliating framework, the question is raised whether fund managers are able to exploit market inefficiencies. Combining time-varying measures of efficiency and performance, we find a positive relationship between α and weak form market efficiency. Most funds are unable to systematically outperform the market, although a few funds do seem to handle relatively inefficient markets well. Top performing funds are characterized by a better management of downside risk in times of market distress, whilst simultaneously exploiting learning effects when markets return to equilibrium. Conditioning fund performance on the state of the underlying market we construct a conditional alpha ratio, which helps in better understanding fund performance and can further improve the fund selection process for investors.

3.1 Introduction

An academic debate on market efficiency has been going on for 50 years. Two opposing convictions are dominating the literature: the efficient market hypothesis (EMH) versus behavioral finance. Under the EMH, an efficient market is defined as "a market in which prices always fully reflect available information" (Fama, 1970, p. 383). Weak form efficient markets fully reflect past price information, semi-strong efficient markets fully reflect publicly available information and strong form efficient markets fully reflect all available information, both public and private. Most proponents of the EMH argue that financial markets are perfectly capable of aggregating public information correctly into the market price, hence turning markets semi-strong form efficient. Alternatively, most opponents of efficient market theory have gathered their evidence and started the field of behavioral finance in close collaboration with psychologists. Behaviorists argue that human beings are not always rational optimizers, which leads to public information being erroneously translated into prices at some times (e.g. over- or under reaction). Both schools of thought have produced numerous empirical research, although not sufficiently convincing to settle the debate. Note that non-behavioral critiques of the EMH also have been formulated (e.g. Lambert, Leuz, & Verrecchia, 2012). For a more detailed overview of the market efficiency literature, we refer to several excellent review papers (Dimson & Mussavian, 1998; Lim & Brooks, 2011; Sewell, 2011; Verheyden, De Moor, & Van den Bossche, 2013).

The ongoing stalemate in empirical finance has been confirmed by the joint awarding of the 2013 Nobel Memorial Prize in Economics to scholars representing the two different sides of the debate. Amongst other significant contributions, Fama is the founding father of the EMH. Summarizing empirical research in seminal review papers (Fama, 1970, 1991, 1998), he still strongly supports the EMH and refutes most arguments of behaviorists. Shiller is a leading scholar from the field of behavioral finance, most renowned for correctly calling asset and housing pricing bubbles (e.g. Shiller, 2000). Even during their respective Nobel Prize lectures, Fama and Shiller openly contradicted each other, further illustrating the apparent irreconcilable nature of empirical research on efficient markets.

As a result of the enduring struggle between proponents of the EMH and behavioral finance, some scholars started to look for alternative theoretical frameworks that could combine the strengths from both schools of thought. The most successful attempt was delivered by Lo (2004, 2005), who established the adaptive markets hypothesis (AMH). Following the notion of time-varying efficiency (Campbell, Lo, & MacKinlay, 1997) and biological evolution, Lo (2004, 2005) states that markets can be fairly efficient for a long period of time, but with behavioral biases causing temporary disturbances. Much like in the field of biology, markets represent an ecology that can exist in equilibrium until a certain shock triggers the process of natural selection and competition, causing the market ecology to change. During a learning and adaptation period, investors compete amongst each other until a new market equilibrium is found and market efficiency returns to pre-shock levels. The 2008 financial crisis with the failing of Lehman Brothers is an illustrative example of a market shock causing the market ecology to change and efficiency levels to be temporarily disturbed. Combining the notion of efficient markets and the impact of behavioral biases in times of market mania, the AMH presents a reconciling theoretical framework that could be interesting to help settle the ongoing debate.

Although promising, the theory of the AMH has not really broken into mainstream finance yet, which is mainly due to a limited amount of supporting empirical evidence. Lo (2004, 2005) confirms the idea of dynamic market efficiency using rolling first-order autocorrelations. The same cyclical efficiency pattern is found on Asia-Pacific financial markets (Todea, Ulici, & Silaghi, 2009) and foreign exchange markets (Neely, Weller, & Ulrich, 2009). Kim, Shamsuddin and Lim (2011) also find weak-form market efficiency to be driven by changing market conditions, although the theoretically predicted pattern of efficiency — with long periods of relative efficiency being alternated with short periods of relative inefficiency — is discarded. Urquhart and Hudson (2013) further confirm that the AMH better describes global financial markets than the EMH. Using different state-of-the-art rolling efficiency tests on developed markets from across the globe (U.S., Europe and Japan), Verheyden, Van den Bossche and De Moor (2014) corroborate both the dynamic character and the predicted pattern of weak-form market efficiency. However, the evidence is most compelling for linear tests on the U.S. stock market, and more limited in scope for the last 15 years on the European and Japanese stock market.

The empirical finding of periods of relative weak form market inefficiency leads to another interesting question in asset pricing: are fund managers able to exploit periods in which the weak form of the EMH does not hold? Under the EMH, fund managers are not believed to be able to generate significant risk-adjusted excess returns (alpha) because of the efficient incorporation of information in asset prices. Starting from seminal work by Jensen (1968), most empirical research confirms that actively managed funds on average underperform passive alternatives (e.g. Carhart, 1997), although it might not be completely irrational to invest in open ended mutual funds (Gruber, 1996). Decomposing mutual fund returns several scholars have shown that fund managers indeed possess significant stock-picking ability, but not sufficiently to overcome the overall mutual fund cost structure (e.g. Wermers, 2000). The central contribution of our paper is the combination of market efficiency and performance evaluation research by examining mutual fund performance in relation with weak form market efficiency. We combine a time-varying weak form efficiency proxy with a time-varying measure for alpha from an unconditional four-factor model (Carhart, 1997). Doing so, we examine the relationship between market efficiency and alpha, and when and how performance is realized by fund managers, conditional on the relative market efficiency of the market. We confirm that most fund managers are not able to outperform the market and market inefficiencies typically lead to lower performances. Only a limited number of fund managers are able to systematically outperform the market and they typically do so by limiting losses in times of market inefficiency and by profiting from subsequent learning effects, consistent with the theory of AMH. Conditioning fund performance on market efficiency, we are able to construct a conditional alpha ratio, which provides a new tool for investors to help identify the best mutual funds. All in all, market efficiency clearly helps in identifying good fund management, which can help investors make better fund selection decisions. Our results also illustrate the possible reconciliation on two levels. First, it provides the insight that efficient market theory and behavioral finance can go hand in hand. Second, it aligns finance academics and professionals, as it shows that active fund management is not always completely obsolete, although markets are generally rather weak form efficient.

The remainder of this paper is organized as follows. In Section 3.2 the data is presented. Section 3.3 provides an overview of both the applied market efficiency and performance evaluation methodology. Results from both methodologies are presented in Section 3.4. The development of a market efficiency perspective on fund performance, combining both methodologies, is given in Section 3.5. Section 3.6 concludes.

3.2 Data

Our methodological approach consists of two main steps. First, we compute a time-varying measure of weak form market efficiency. Next, we calculate a time-varying measure for alpha from an unconditional four-factor model. We then link the rolling weak form efficiency measure to the rolling alpha to obtain more insights on our central research question. In Section 3.2 and 3.3 we present the data and the methodology along the lines of these two steps.

3.2.1 Weak form market efficiency data

To obtain a robust time-varying measure of weak form market efficiency, we use daily return data organized in rolling windows of one year (260 observations), following recommendations on weak form efficiency tests from Verheyden et al. (2014). The focus of our research is on the U.S. stock market, which we represent in the same way as Fama and French (1993), i.e. "by all CRSP firms incorporated in the U.S. and listed on the NYSE, AMEX, or NASDAQ that have a CRSP share code of 10 or 11 at the beginning of month t, good shares and price data at the beginning of t, and good return data for t minus the one-month Treasury bill rate (from Ibbotson Associates)." Daily return data on the U.S. market proxy is taken from the website of Kenneth French. In comparing the time-varying efficiency measure with the unconditional four-factor rolling alpha, we take corresponding time periods. As we want to demonstrate the usefulness of the market efficiency concept as a way of better understanding superior fund management ability, we focus on a sample of data from the last ten years, which is set around the most recent financial crisis[1]. The evolution of the daily returns is plotted in Figure 3.1. Summary statistics are presented in Table 3.1.

[1] Note that market inefficiencies also occurred outside of the most recent financial crisis (e.g. more than 902 consecutive days between 1983 and 1987 and more than 351 consecutive days between 1994 and 1996). Our results are hence also insightful outside of times of extreme market conditions like during the 2008–2009 financial crisis.

Summary statistics: Daily returns	
Begin	2004/07/01
End	2014/06/30
Mean	0.040200
Median	0.095000
Minimum	-8.9500
Maximum	11.354
Standard deviation	1.2880
Coefficient of variation	32.039
Skewness	-0.14187
Excess kurtosis	9.9357
5% percentile	-1.9227
95% percentile	1.7310
Interquartile range	1.0590

Table 3.1: *Summary statistics*

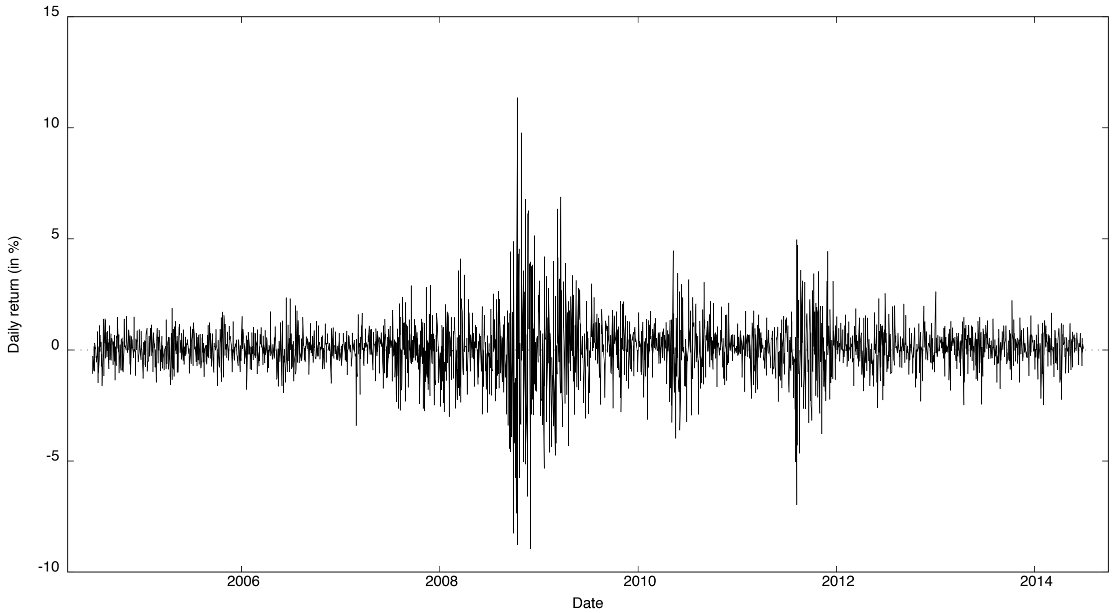


Figure 3.1: *Evolution of daily returns*

3.2.2 Mutual fund performance data

As we measure the weak form market efficiency of the U.S. stock market, we focus on U.S. equity mutual funds. The central goal of our paper is to compare the performance of fund managers with respect to the underlying levels of weak form market efficiency. In order to truly capture time-varying performance of a fund manager, we opt for a well-motivated selection of U.S. equity mutual funds, rather than pursuing a fund-of-funds strategy including the entire set of listed U.S. equity mutual funds. We specifically look for all U.S. equity mutual funds for which the average portfolio manager tenure is at least ten years. Requiring a fund manager tenure of at least ten years allows us to better capture and attribute the ability of the fund manager to outperform in turbulent markets over a reasonably long period and avoids noisy performance measurement. Including all funds with complete data for our time frame and an average manager tenure of at least ten years, and excluding any remaining index/passively managed funds, we arrive at a sample of 272 U.S. equity mutual funds (more details see Appendix 3.A). Note that the issue of survivorship bias is not really relevant in our setting, as we are not looking at historical outperformance of funds, but rather analyze the relationship between market efficiency and fund performance for live funds. The proposed methodology in this paper (a way to identify a manager's ability to exploit market inefficiencies) is meant to be helpful in the fund selection decision of investors, irrespective of their initial pool of funds. In this selection process, dead funds are not considered. Also, note that in the process of closing down a fund it takes some time to settle outstanding positions, which might bias results upon inclusion.

For our sample of 272 U.S. equity mutual funds, we retrieve daily total net returns from Morningstar, controlling for management fees. Corresponding daily data on the four included factors of risk are collected from the website of Kenneth French, following the Fama and French (1996) methodology.

3.3 Methodology

3.3.1 Test of time-varying weak form market efficiency

Weak form market efficiency in itself cannot be observed from the market. Hence, researchers typically rely on a proxy measure for efficiency. An overview of different proxy alternatives is presented in Chapter 1. An established way to test for weak form market efficiency is by means of the return predictability proxy. If market prices cannot be predicted from historical price information, the market is deemed weak form efficient and vice versa. Testing for return predictability is possible using variance ratio (VR) tests, which were first developed by Lo and MacKinlay (1988) as a test for an uncorrelated increment. If stocks follow an uncorrelated increment or random walk, they are believed to be unpredictable from prior price information. Consequently, technical trading rules should be ineffective, which is then being interpreted as a proxy for weak form market efficiency. It is important to keep this line of reasoning in mind and realize that this only constitutes a proxy measure for the concept of weak form market efficiency. An uncorrelated increment occurs when the variance of the stock returns over a certain interval k are statistically indistinguishable from k times the variance of these stock returns over a time interval of 1. This principle can easily be translated into a VR test statistic (Lo and MacKinlay, 1988):

$$VR(k) \equiv \frac{Var(r_t(k))}{kVar(r_t)} = 1 + 2\sum_{j=1}^{k-1}(1 - \frac{j}{k})\rho_j \tag{3.1}$$

With r_t the daily return on the U.S. stock market on day t ($t = 1, \cdots, T$), $r_t(k) \equiv r_t + r_{t-1} + \cdots + r_{t-k+1}$, and ρ_j the j^{th} order autocorrelation of r_t. The choice of the vector of holding periods k is arbitrary, although Deo and Richardson (2003) advocate the use of holding periods of 2, 5 and

10 days. The null hypothesis associated with this test statistic assumes that the return series are following an uncorrelated increment. If the test statistic is significantly different from unity, the null is rejected, which can be interpreted as the return series not being weak form market efficient.

Ever since this first attempt, many improvements have been made to the original VR test statistic. Choi (1999) transformed the original VR test to an automatic VR (AVR) test such that the choice of the vector of holding periods k was determined by a fully data-dependent procedure, following Andrews (1991) using an asymptotic truncated mean squared error criterion, instead of an arbitrary decision.

$$AVR(\hat{k}) = \sqrt{\frac{T}{k}} \frac{[VR(\hat{k}) - 1]}{\sqrt{2}} \xrightarrow{d} N(0,1) \tag{3.2}$$

The AVR test statistic converges to a standard normal distribution when the sample size approaches infinity, but it is unknown what happens when the sample size is small. Consequently, the use of an asymptotic approximation for critical values can lead to a serious size distortion. A bootstrap method provides a resampling alternative that avoids reliance on large sample theory. More specifically, a wild bootstrap version of this AVR test statistic, following the wild bootstrap procedure of Mammen (1993) for time series that exhibit conditional heteroscedasticity, is currently considered state-of-the-art to test for weak form market efficiency (Kim, 2009; Charles, Darné, & Kim, 2011). The wild bootstrapping procedure takes into account possible heterogeneity in the daily returns and yields a heteroscedasticity-consistent confidence interval for the AVR test statistic. This test is also shown to be more powerful (higher probability of correctly rejecting the null hypothesis) than alternative tests presented in the previous chapter. We use 500 iterations to perform the test.

To obtain a time-varying measure of weak form market efficiency, we implement a rolling version of the wild bootstrap AVR test. We estimate the rolling wild bootstrap AVR test statistic for corresponding periods of U.S.

stock market and fund data, along with the associated 95% confidence interval. In a first window, we include the first 260 return observations and execute our estimation approach, after which the window is pushed forward by one observation. Continuing this procedure until the last observation is included in the rolling sample for the first time, we obtain a time-varying measure of return predictability (i.e. autocorrelation), together with the rolling 95% confidence interval, which can then be interpreted as a proxy for time-varying weak form market efficiency. Note that whenever the test statistic falls outside of the confidence interval, the U.S. market can be interpreted as not being weak form efficient. We implement the test in R using the *vrtest* package.

3.3.2 Test of time-varying mutual fund performance

We implement an unconditional four-factor model to compute risk adjusted returns (alpha). Following Fama and French (1996) and Carhart (1997) we control for four factors of risk: market risk ($r_{Mt} - r_{Ft}$), size risk (SMB_t), book-to-market risk (HML_t) and momentum risk (MOM_t).

$$r_t - r_{Ft} = \alpha + \beta_1(r_{Mt} - r_{Ft}) + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 MOM_t + \varepsilon_t \quad (3.3)$$

With r_t the return of the U.S. equity mutual fund on day t; r_{Ft} the risk-free return approximated by the one-month T-bill rate on day t; ($r_{Mt} - r_{Ft}$) the excess return on the market portfolio on day t; SMB_t the daily return on a zero-investment difference portfolio that is long in a small-cap portfolio and short in a large-cap portfolio; HML_t the daily return on a zero-investment difference portfolio that is long in a value-type portfolio and short in a growth-type portfolio; and MOM_t the daily return on a zero-investment difference portfolio that is long in a portfolio with winning stocks and short in a portfolio of losing stocks based on the past year. Assuming that this is the true underlying asset pricing model and hence that all relevant factors of risk have been taken into account and that the U.S. market benchmark is an appropriate proxy of the market portfolio, α represents the risk-adjusted excess return.

This methodology is implemented in rolling windows of one year (260 observations)[2], starting from the first return observation on the first of July, 2004. As a result, we find a time-varying measure of the risk-adjusted excess return (α), which can now be compared with the matched proxy of time-varying weak form market efficiency for a corresponding period of returns. Next to the alpha estimate, we obtain a 95% confidence interval using heteroscedasticity and autocorrelation consistent (HAC) standard errors. Note that our use of rolling-windows to estimate time-varying mutual fund performance is well-motivated. As we want to gain more insights into the ability of fund managers to exploit market inefficiencies, we mimic the best-established time-varying return predictability test, which draws from a rolling-window strategy. This enables us to efficiently compare both time-varying measures in a straightforward way. More sophisticated estimation strategies like time-varying parameter models are also available, but it would become much more tedious to jointly present and analyze the obtained time-varying measures of weak form efficiency and alpha.

Our proposed methodological strategy yields a corresponding sample of time-varying measures for weak form market efficiency and alpha for each of the 272 funds under study. From these corresponding results, we can now look into the ability of fund managers to generate risk-adjusted excess returns, and to what extent this ability is driven by periods when the weak form of the EMH is rejected.

[2] Just like for the time-varying variance ratio test, we tested for the robustness of the window length and confirm that a window of 260 observations leads to reliable results.

3.4 Results

3.4.1 Time-varying weak form market efficiency

We collect the wild bootstrap AVR test statistics and the 95% critical values (in grey) for every window and plot the results in Figure 3.2. Note that the date for every window is taken as the average date between the beginning and the end of the window. The obtained time-series of test statistics represents a measure of time-varying return autocorrelation, which can in turn be interpreted as a proxy measure for time-varying weak form market efficiency. Whenever the test statistic falls outside of the 95% critical range, we reject the null hypothesis of a weak form efficient market. From Figure 3.2, we see that the weak form market efficiency of the U.S. market is not constant and changes over time, as predicted by the AMH (Lo, 2004, 2005). Note that the most recent manifestation of market inefficiency occurred during the 2008-2009 market turmoil.

Next to a graphical analysis, we can also summarize the efficiency results by calculating an efficiency ratio (ER), which can be interpreted as the relative degree of weak form market efficiency of the underlying market. The ER is defined as follows.

$$ER = 1 - \frac{\#\ windows\ rejecting\ null\ of\ no\ return\ predictability}{Total\ \#\ windows} \tag{3.4}$$

Over the period of 2004/07/01 to 2014/06/30 we find an ER of 79.72% for the U.S. market, using the wild bootstrap AVR test. Hence the U.S. stock market is considered weak form market efficient 79.72% of the time over this ten year period. Alternatively, this means that 20.28% of the time the market is not considered weak form efficient and thus could be exploited to generate risk-adjusted excess returns (alpha). We also list the periods characterized by significant return predictability, and thus relative weak

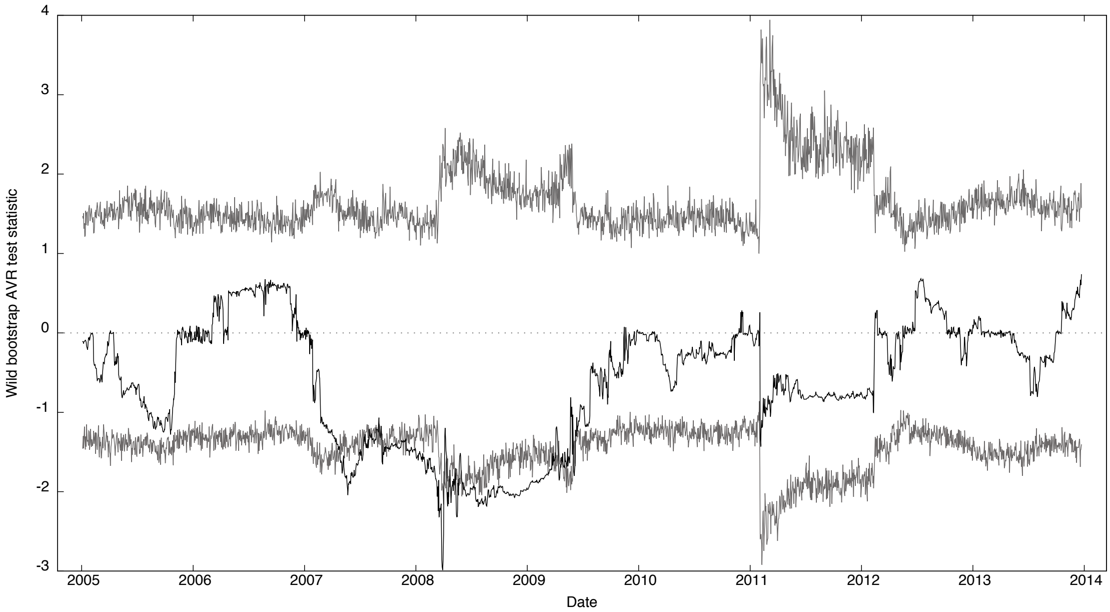


Figure 3.2: *Rolling return autocorrelation of the U.S. stock market*

form market inefficiency (Table 3.2). Note that we only include prolonged periods where the null hypothesis of no return predictability is rejected for 40 (two months) or more days in a row. This allows us to detect periods with significantly lower degrees of weak form market efficiency and sustained deviations from the efficient market equilibrium, rather than infrequent rejections of the null hypothesis which might reflect statistical flukes. These periods of prolonged market inefficiency are of particular interest to further examine the ability of fund managers to generate alpha.

Inefficiency periods		Consecutive days
2007/04/27	2007/07/12	53
2007/12/01	2008/04/02	86
2008/07/11	2009/03/09	167

Table 3.2: *Prolonged periods of relative weak form market inefficiency*

Despite evolutions in communication technology over the last 50 years, we cannot conclude that the U.S. market has become immune to weak form market inefficiencies, as proven by the more than 150 consecutive days with significant return predictability during the most recent financial crisis. Given that these inefficiencies already started in early 2007, confirming the what was then called irrational behavior, one could even argue that the 2008 financial meltdown could have been foreseen from this time-varying measure of weak form market efficiency. All in all, the results confirm the idea of time-varying weak form market efficiency, as suggested by Campbell et al. (1997) and Lo (2004, 2005).

We can also project these periods of prolonged relative weak form market efficiency (in grey) on an index for the broad U.S. market (Figure 3.3).

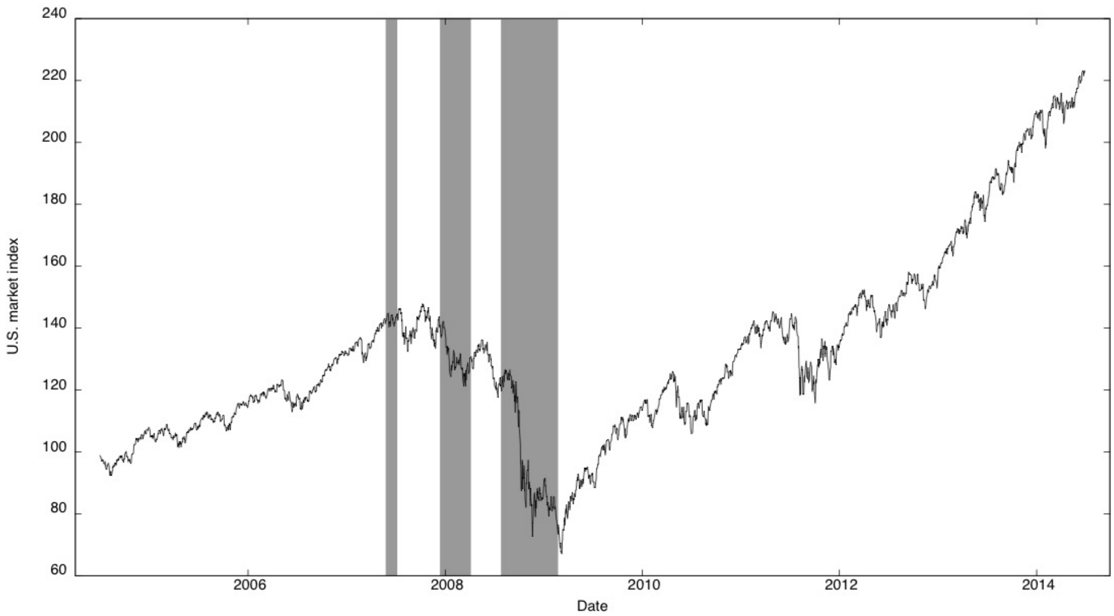


Figure 3.3: *Time series of the U.S. market index*

From the graph, we see that the prolonged periods of relative market inefficiency coincide with the times of market distress and the largest price drop during the most recent financial crisis, which is in line with our earlier presentation of the market efficiency concept.

3.4.2 Time-varying risk-adjusted excess returns

From our proposed estimation strategy, we find time-varying measures of a four-factor unconditional alpha for each of the 272 U.S. equity mutual funds in our sample. The α time series for the full sample is displayed in Figure 3.4.

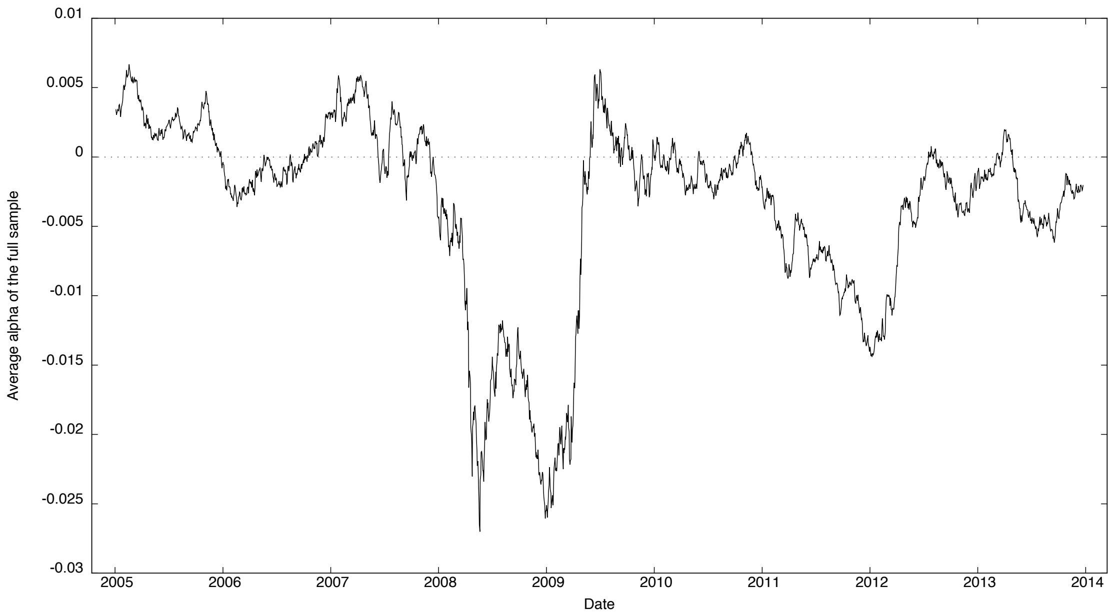


Figure 3.4: *Time series of α*

Having a look at the time series, we see that α tends to vary heavily through time. Periods of declining α seem to coincide with periods of increased market distress. To summarize these results we can also calculate a so-called positive (PAR) and negative alpha ratio (NAR), which we define as follows.

$$PAR = \frac{\#\ windows\ significant\ positive\ \alpha}{Total\ \#\ windows} \tag{3.5}$$

$$NAR = \frac{\#\ windows\ significant\ negative\ \alpha}{Total\ \#\ windows} \tag{3.6}$$

The summary statistics for the alpha ratios are presented in Table 3.3; the box plot is displayed in Figure 3.5.

Summary statistics

	Positive alpha ratio	*Negative alpha ratio*
Mean	0.024895	0.053496
Median	0.0068645	0.030337
Minimum	0.0000	0.0000
Maximum	0.21789	0.31754
Standard deviation	0.036685	0.063645
Coefficient of variation	1.4683	1.1897
Skewness	1.9648	1.6491
Excess kurtosis	4.4080	2.8319
5% percentile	0.0000	0.0000
95% percentile	0.10089	0.18279
Interquartile range	0.041519	0.074734

Table 3.3: *Summary statistics of the α ratio*

From the summary statistics and box plot we observe that α performance varies heavily across the different funds. Most funds seem unable to systematically generate positive alpha, whereas more funds generate negative alpha over our ten-year sample (higher mean, median and maximum value). From both alpha ratios we can already obtain a first look into the performance profile of a fund manager. Ideally, a fund would achieve a high positive alpha ratio, whilst simultaneously limiting its negative alpha ratio. For example, the Wasatch Micro Cap Value Fund achieves the highest positive alpha ratio (21.8%), whilst limiting its negative alpha ratio to

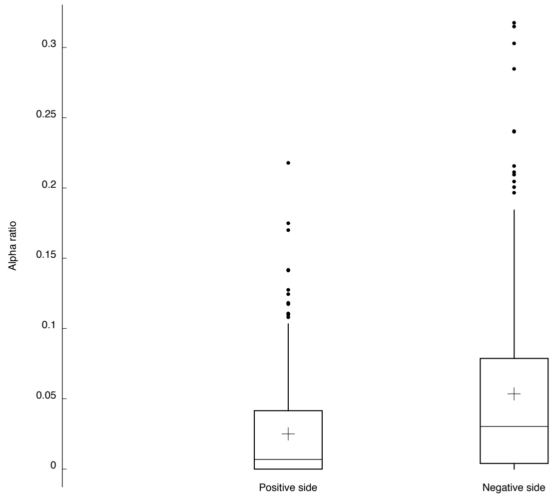


Figure 3.5: *Box plot of the α ratios*

7.5%. On the other side of the spectrum, the STAAR Larger Company Stock Fund generates a negative alpha 31.8% of the time, while generating virtually no significant positive alpha. Apart from the top- and bottom performers, 14 funds achieve a zero ratio for both the positive and negative side of alpha, which could be a sign of closet indexing. In general there is only limited correlation and small rank correlation between the positive and negative side of the alpha ratio, showing that funds that achieve significant positive α typically limit the amount of downside risk.

The use of a time-varying measure for α and the calculation of alpha ratios clearly adds understanding to the performance of mutual funds, and may help fund investors to make better decisions. To further capture a fund manager's ability to deal with market crises, we can also take into account the underlying level of weak form market efficiency in the next section.

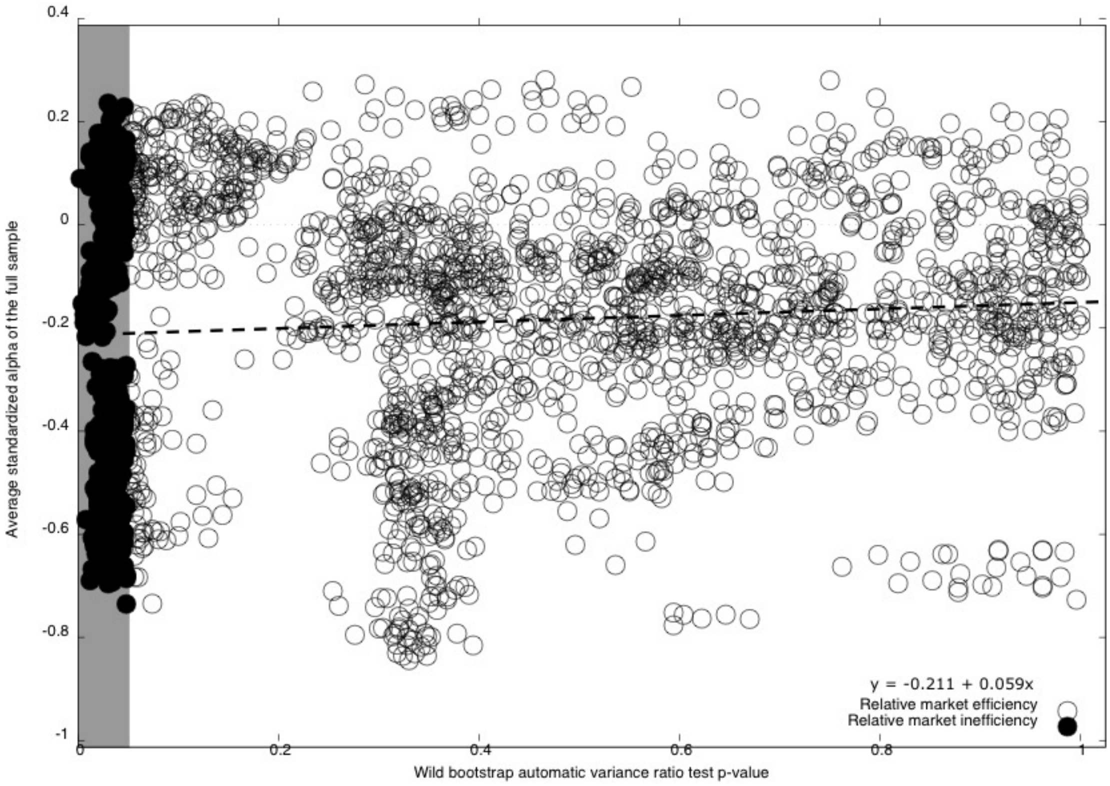
3.5 Market efficiency perspective on fund performance

After computing time-varying measures for α and weak form market efficiency separately, we now want to combine both methodologies to help better understand fund performance. Following our central research question, we want to find out if there is a relationship between mutual fund performance and the underlying state of the market. Therefore, we start by plotting our proxy for weak form market efficiency with respect to the standardized α obtained in the corresponding window for our full sample of 272 U.S. equity mutual funds (Figure 3.6).

Plotting market efficiency versus α creates an interesting picture of mutual fund performance. We observe a positive relationship between the wild bootstrap automatic variance ratio test p-value and α, which indicates that fund performance improves when the underlying market is becoming relatively more efficient. This need not be surprising, as market inefficiencies typically coincide with increased market turmoil. Also note that performance in general is quite poor, but especially so in times of relative market inefficiency, which also becomes clear from the box plot.

Looking at the box plot (Figure 3.7), we note that the interquartile distance in performance across funds is larger in times of relatively inefficient markets, which means that it might be easier to distinguish good from bad fund managers using the market efficiency perspective. We further look into this property calculating conditional alpha ratios. Having combined both methodologies, we can also revisit Figure 3.4, by adding the prolonged periods of relative market inefficiency in grey (Figure 3.8).

We now observe the impact of relative market inefficiency on the average α for the full sample. The largest drop in aggregate α indeed occurs in a time of relative market inefficiency, which coincides with the market turmoil caused by the collapse of Lehman Brothers in September of 2008.



As a proxy for market efficiency we take the p-value of the wild bootstrap automatic variance ratio test, with the null hypothesis that the underlying market is weak form efficient. In the shaded area, this null hypothesis is rejected with 95% confidence. All observations when the null hypothesis was rejected are indicated in full black circles; the other observations are in white circles. We also plot the regression relationship between both variables.

Figure 3.6: *Scatter plot of the proxy of market efficiency versus the obtained standardized alpha in the corresponding window*

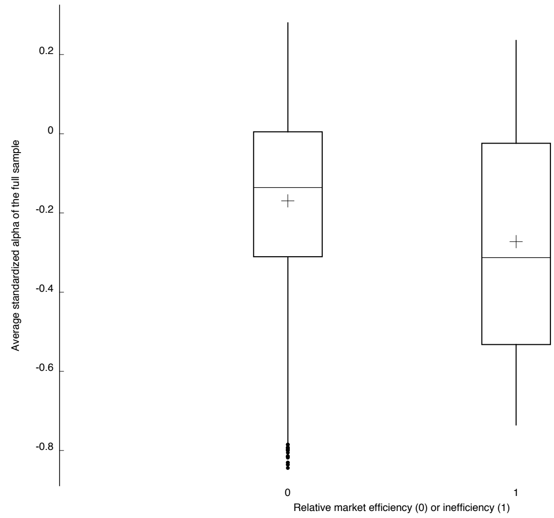


Figure 3.7: *Box plot of the standardized average α conditional on the efficiency of the underlying market*

After the third and last prolonged period of relative market inefficiency, the market returned to equilibrium and average fund performance improved.

We are also interested in the differences between funds that generate a significant positive (6) or negative alpha (14), and funds that do not generate any significant alpha (252) over the entire ten-year period. We plot the different times series along with periods of prolonged relative market inefficiency (in grey) in Figure 3.9.

From the breakdown, we obtain a further insight into the link between market efficiency and fund management skills. The six funds that generate a significantly positive alpha over the entire period seem better able to limit the drawdown in times of market inefficiency, whilst still fully profiting from subsequent learning effects when the market returns to equilibrium. Even in times of market turmoil, the positive alpha group is able to maintain fairly good returns, whilst the other two groups are suffering losses and take a longer time to recover.

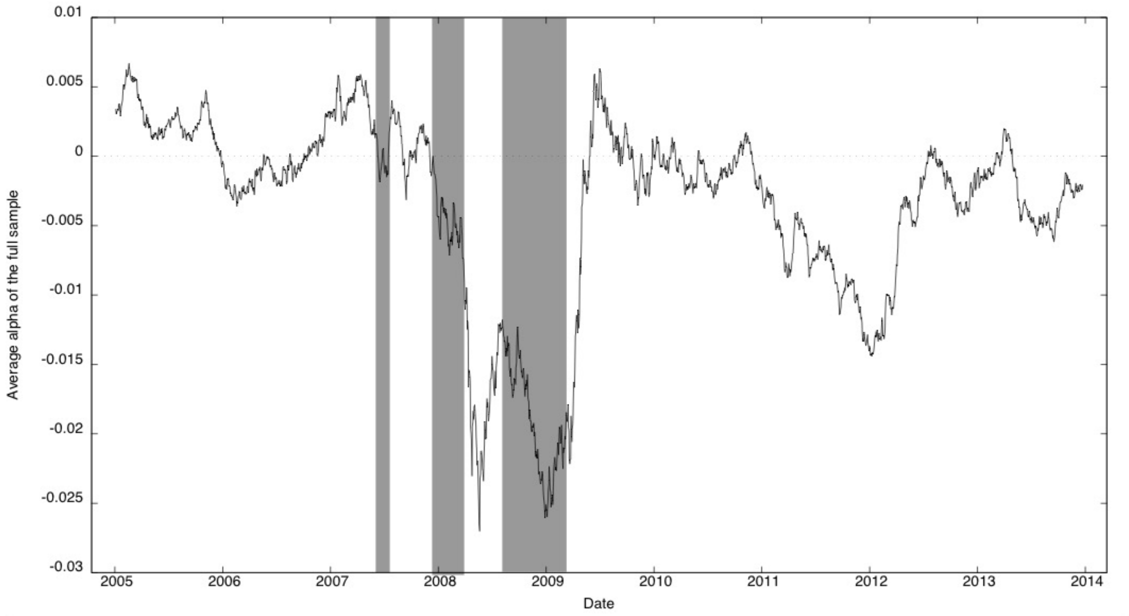


Figure 3.8: *Time series of α (bis)*

As a general takeaway from our graphical analyses, we learn that there can be substantial differences between the performance of mutual funds, and that these differences can be better understood from a market efficiency perspective. Consequently, it can be insightful to create a further adaptation of the alpha ratios to better help understand mutual fund performance. Instead of looking at the creation of significant positive and negative alpha in the overall market, we now condition on periods of relative market inefficiency. We dub this the conditional alpha ratio, which again has a positive (PCAR) and negative (NCAR) side.

$$PCAR = \frac{(\#\ windows\ significant\ positive\ \alpha\ \mid inefficiency)}{Total\ \#\ inefficient\ windows} \tag{3.7}$$

$$NCAR = \frac{(\#\ windows\ significant\ negative\ \alpha\ \mid inefficiency)}{Total\ \#\ inefficient\ windows} \tag{3.8}$$

The summary statistics for the conditional alpha ratio are presented in

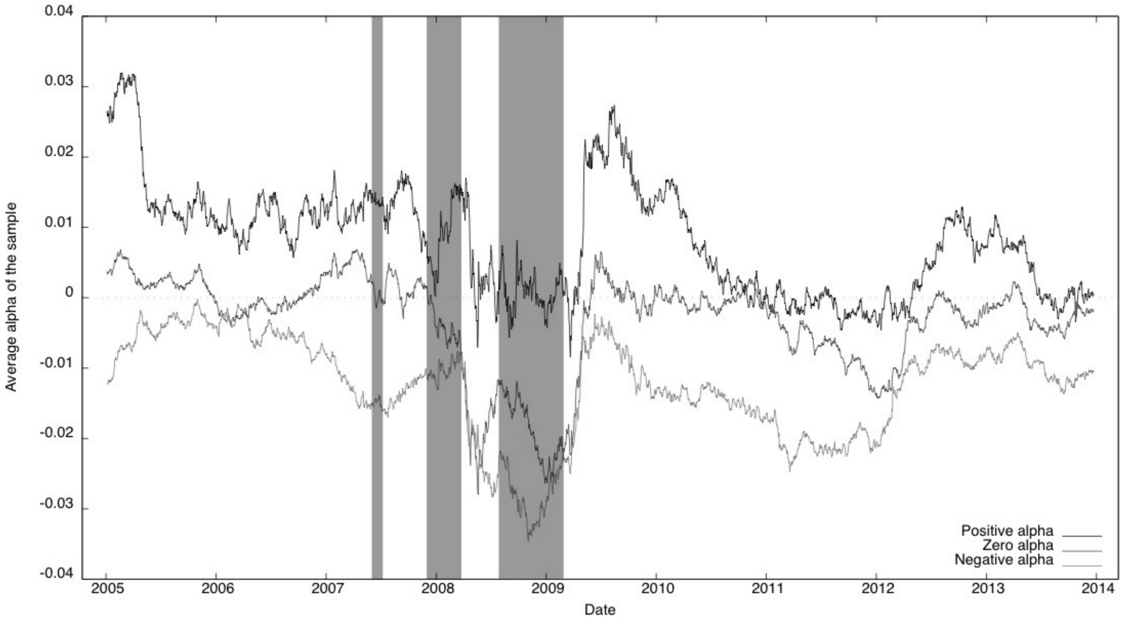


Figure 3.9: *Time series of α for the breakdown of positive, negative and zero α funds*

Table 3.4. Figure 3.10 presents the box plot of both sides of the conditional alpha ratio.

Both the summary statistics and the box plot show a large disparity between funds, indicating once more that the market efficiency perspective facilitates the fund comparison/selection process. 102 out of 272 mutual funds in our sample simultaneously have a zero PCAR and NCAR, which means that they neither out- nor underperform the overall market in times of distress. This also leads to the median values of both sides being zero. Apart from this finding, however, we also find a great number of “outliers”, i.e. funds that either do very well or very bad according to the conditional alpha ratio. On the bottom-end, the Royce Premier Consult Fund generates a negative alpha 61.4% of the time when the market is relatively weak form inefficient, while at the seem time never generating significant positive alpha in the same market conditions. Numerous other funds are

Summary statistics

	Positive conditional alpha ratio	Negative conditional alpha ratio
Mean	0.018639	0.051085
Median	0.0000	0.0000
Minimum	0.0000	0.0000
Maximum	0.33188	0.61354
Standard deviation	0.057732	0.10426
Coefficient of variation	3.0974	2.0410
Skewness	3.7538	2.7362
Excess kurtosis	13.929	7.5636
5% percentile	0.0000	0.0000
95% percentile	0.15895	0.32544
Interquartile range	0.0000	0.049672

Table 3.4: *Summary statistics of the conditional α ratio*

characterized by the same kind of performance. On the top end, there are a couple of funds that manage to achieve a decent PCAR, whilst keeping the downside risk to a minimum. For example, the Columbia Small Cap Value Fund II generates significant positive alpha in 30.3% of the windows where market efficiency is rejected, without any negative alpha in the same conditions. The correlation between the positive and negative side of the conditional alpha ratio is limited and there is no rank correlation. Fund managers who are able to generate positive significant α in times of market distress, typically also manage to limit the drawdown under the same circumstances.

Comparing the alpha and conditional alpha ratio, there is a good amount of (rank) correlation, as both measures help characterize fund performance. However, the conditioning of performance on the underlying state of the market further adds to the process of distinguishing good from bad funds, as shown by the complete lack of rank correlation between conditional alpha ratios. This further confirms the added value of the market efficiency

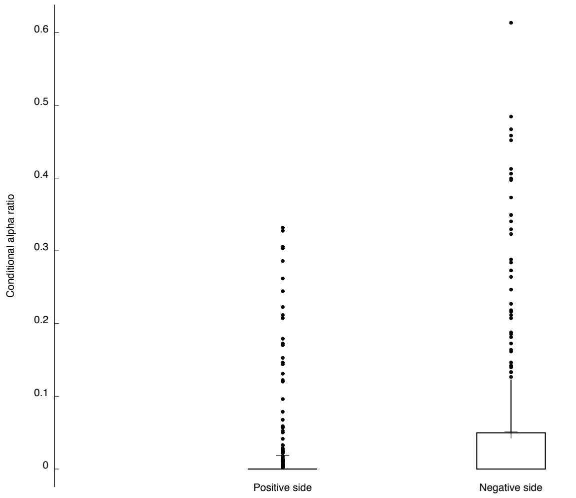


Figure 3.10: *Box plot of the conditional α ratios*

perspective. All in all, combining performance and efficiency measures can help investors to better understand fund performance and the ability of a fund manager to simultaneously outperform the market whilst limiting downside risk in times of market distress. The conditional alpha ratio can hence be a useful decision tool in picking mutual funds, along with more traditional criteria that can be taken into account.

3.6 Conclusion

Implementing a rolling-window four-factor performance model, we find the ability to create α to be time-varying and different between funds. A corresponding rolling proxy for weak form market efficiency confirms the general idea behind the adaptive markets hypothesis (Lo, 2004, 2005), i.e. market efficiency is a relative concept that can change through time because of temporary market distress and deviations from equilibrium. The central contribution of our paper is the combination of both time varying measures to help better understand mutual fund performance. We find a positive relationship between α and efficiency, meaning that funds typically perform better in times of market efficiency. Generally, mutual funds have a very hard time generating α in a systematic way, as was found in earlier research as well. However, adding the market efficiency dimension to our analysis accommodates the process of identifying those funds that actually do create value for an investor through active management. The few funds that are able to outsmart the market are characterized by a better management of drawdown in times of market distress, whilst maximizing the learning effects when markets subsequently return to equilibrium. Conditioning fund performance on the state of the underlying market clearly helps in better understanding fund performance. The conditional alpha ratios can be instrumental for investors to improve their fund selection process.

Our work also represents two important insights in the academic debate on market efficiency and active fund performance. First, the use of time-varying weak form efficiency measures reconciles the extreme views of the EMH and behavioral finance. As predicted by the AMH (Lo, 2004, 2005) and found by earlier empirical research, we confirm that weak form market efficiency is not constant over time but rather changes with underlying market conditions. Second, the application of the time-varying efficiency construct to decompose active mutual fund performance shows that top active fund managers are not entirely unable to outperform the market, as is argued by advocates of the EMH and is found in most earlier research. Given that markets can temporarily become inefficient, good active fund managers can exploit these market inefficiencies to generate significant α.

Our current research can be strengthened in a number of ways. The (conditional) alpha ratios could be tested for their ability to forecast future performance and tail risks of funds. Additionally, a detailed analysis of dead funds in relationship with market efficiency could further add to our understanding of the adaptive markets hypothesis.

Appendix 3.A: List of funds in data sample

#	ISIN	Name	Inception date	Global category	Fund size (in $)	Annual net expense ratio
1	US8162211051	Selected American Shares S	28/02/33	US Equity Large Cap Blend	5,812,176,427	0.94
2	US0454198013	AMF Large Cap Equity AMF	30/06/53	US Equity Large Cap Blend	62,171,327	1.22
3	US56064V2051	Mairs & Power Growth Inv	7/11/58	US Equity Large Cap Blend	4,251,925,094	0.67
4	US8568391051	State Farm Growth	1/05/67	US Equity Large Cap Blend	4,135,700,000	0.12
5	US94904P6097	Weitz Partners Value?Investor	31/05/83	US Equity Large Cap Blend	1,075,015,197	1.18
6	US2619701079	Dreyfus Appreciation Investor	18/01/84	US Equity Large Cap Blend	5,935,346,731	0.94
7	US94904P2039	Weitz Value?Investor Class	9/05/86	US Equity Large Cap Blend	1,144,259,608	1.18
8	US0228651099	Amana Income Investor	23/06/86	US Equity Large Cap Blend	1,636,413,353	1.18
9	US8888941023	Tocqueville	13/01/87	US Equity Large Cap Blend	388,033,351	1.26
10	US74440N2018	Prudential Jennison Value B	22/01/87	US Equity Large Cap Blend	718,024,609	1.79
11	US5430691080	Longleaf Partners	8/04/87	US Equity Large Cap Blend	8,595,961,136	0.92
12	US20651N1090	Concorde Value	4/12/87	US Equity Large Cap Blend	12,263,888	2.09
13	US9620961030	Muhlenkamp	1/11/88	US Equity Large Cap Blend	487,237,287	1.26
14	US8914021097	Torray	18/12/90	US Equity Large Cap Blend	418,630,342	1.15
15	US8848913005	Thompson LargeCap	10/02/92	US Equity Large Cap Blend	127,037,711	1.21
16	US00170K5882	AMG Yacktman Service	6/07/92	US Equity Large Cap Blend	14,182,838,870	0.74
17	US52469C2070	ClearBridge All Cap Value B	6/11/92	US Equity Large Cap Blend	2,020,874,745	2.4
18	US9695576028	Jamestown Equity	1/12/92	US Equity Large Cap Blend	31,315,279	1.03
19	US4812A14499	JPMorgan Growth & Income B	4/11/93	US Equity Large Cap Blend	465,745,880	1.66
20	US2270611088	Croft Value R	10/05/95	US Equity Large Cap Blend	101,327,060	1.3
21	US6813831053	Olstein All Cap Value C	21/09/95	US Equity Large Cap Blend	688,183,450	2.3
22	US7236822091	Pioneer B	1/07/96	US Equity Large Cap Blend	5,334,669,273	2.24
23	US74437E7013	Prudential Jennison Equity Opp C	7/11/96	US Equity Large Cap Blend	574,204,264	1.8
24	US6933914011	PIMCO StocksPLUS B	20/01/97	US Equity Large Cap Blend	1,131,625,985	1.65
25	US63872R2022	Natixis CGM Advisor Targeted Equity B	28/02/97	US Equity Large Cap Blend	564,033,609	1.91
26	US00170K5700	AMG Yacktman Focused Service	1/05/97	US Equity Large Cap Blend	11,921,561,805	1.25
27	US1253255061	CGM Focus	3/09/97	US Equity Large Cap Blend	1,483,535,087	1.09
28	US2619783655	Dreyfus Tax-Managed Growth C	4/11/97	US Equity Large Cap Blend	196,547,968	2.1
29	US1032011090	Boyar Value	5/05/98	US Equity Large Cap Blend	22,265,031	1.75
30	US4165286695	Hartford Disciplined Equity HLS IB	29/05/98	US Equity Large Cap Blend	866,402,954	1.01
31	US4268941014	Henssler Equity Investor	10/06/98	US Equity Large Cap Blend	100,762,787	1.27
32	US00758M2614	Cambiar Opportunity Inv	30/06/98	US Equity Large Cap Blend	851,836,924	1.2
33	US5534221065	The MP 63	1/03/99	US Equity Large Cap Blend	53,661,354	0.77
34	US19765H2893	Columbia Select Large Cap Equity Fund C	2/08/99	US Equity Large Cap Blend	536,505,419	1.94
35	US5851051095	Meehan Focus	10/12/99	US Equity Large Cap Blend	57,954,045	1
36	US66538E6480	Marathon Value Portfolio	28/03/00	US Equity Large Cap Blend	68,422,458	1.21
37	US87234N5757	TCW Concentrated Value N	1/03/01	US Equity Large Cap Blend	10,593,803	1.14
38	US4138387075	Oakmark II	5/04/01	US Equity Large Cap Blend	15,275,884,635	1.23
39	US0075W07180	Westwood Dividend Growth Institutional	6/08/01	US Equity Large Cap Blend	83,612,363	0.93
40	US09532K1034	Blue Chip Investor	31/12/01	US Equity Large Cap Blend	24,759,542	1.27
41	US00768D5582	Fort Pitt Capital Total Return	31/12/01	US Equity Large Cap Blend	53,980,364	1.24
42	US4613086036	American Funds Invmt Co of Amer 529B	15/02/02	US Equity Large Cap Blend	74,219,553,376	1.5
43	US9264641575	Victory Diversified Stock C	28/02/02	US Equity Large Cap Blend	1,456,167,891	1.9
44	US2619781345	Dreyfus Core Equity C	15/04/02	US Equity Large Cap Blend	366,728,561	2.1
45	US57629S3498	MassMutual Select Focused Value R3	30/12/02	US Equity Large Cap Blend	930,951,147	1.6
46	US6706903126	Nuveen Large Cap Select C	31/01/03	US Equity Large Cap Blend	49,015,813	2.07
47	US6401931083	Neiman Large Cap Value	1/04/03	US Equity Large Cap Blend	31,520,023	1.45
48	US72200Q5962	PIMCO StocksPLUS Absolute Return C	31/07/03	US Equity Large Cap Blend	1,231,045,975	1.79
49	US36240D4025	Gabelli Dividend Growth C	31/12/03	US Equity Large Cap Blend	35,579,676	2.75
50	US03332V5003	Ancora Equity C	5/01/04	US Equity Large Cap Blend	9,749,437	2.54

#	ISIN	Name	Inception date	Global category	Fund size (in $)	Annual net expense ratio
51	US90470K8181	Crawford Dividend Growth C	26/01/04	US Equity Large Cap Blend	112,161,478	1.89
52	US3786907887	Glenmede Large Cap Core Port	27/02/04	US Equity Large Cap Blend	596,581,068	0.86
53	US3161531054	Fidelity	30/04/30	US Equity Large Cap Growth	5,930,980,248	0.55
54	US2862811005	Elfun Trusts	27/05/35	US Equity Large Cap Growth	2,405,098,903	0.15
55	US0930011053	William Blair Growth N	20/03/46	US Equity Large Cap Growth	920,343,478	1.19
56	US1084391002	Bridges Investment	1/07/63	US Equity Large Cap Growth	119,612,639	0.85
57	US3160711095	Fidelity® Contrafund®	17/05/67	US Equity Large Cap Growth	109,600,198,124	0.66
58	US8174181060	Sequoia	15/07/70	US Equity Large Cap Growth	8,049,462,339	1
59	US6641991061	Northeast Investors Growth	27/10/80	US Equity Large Cap Growth	68,854,171	1.23
60	US9114766044	US Global Investors All American Eq	4/03/81	US Equity Large Cap Growth	23,787,982	2.16
61	US8297971090	Sit Large Cap Growth	2/09/82	US Equity Large Cap Growth	190,477,116	1
62	US3162001040	Fidelity® Growth Company	17/01/83	US Equity Large Cap Growth	43,059,759,940	0.83
63	US2391032032	Davis Opportunity B	1/05/84	US Equity Large Cap Growth	634,854,473	1.95
64	US9219361006	Vanguard PRIMECAP Inv	1/11/84	US Equity Large Cap Growth	43,417,263,302	0.45
65	US7017651099	Parnassus	27/12/84	US Equity Large Cap Growth	631,617,154	0.86
66	US60934G8024	Monetta	6/05/86	US Equity Large Cap Growth	55,463,887	1.51
67	US74405V1070	Provident Trust Strategy	30/12/86	US Equity Large Cap Growth	174,158,077	1
68	US7141993045	Permanent Portfolio Aggressive Growth	2/01/90	US Equity Large Cap Growth	48,686,804	1.2
69	US9695573058	Government Street Equity	18/06/91	US Equity Large Cap Growth	94,383,969	0.84
70	US0228652089	Amana Growth Investor	3/02/94	US Equity Large Cap Growth	2,022,839,880	1.11
71	US7508691091	Rainier Large Cap Equity Original	10/05/94	US Equity Large Cap Growth	507,977,305	1.12
72	US36559B2034	Chesapeake Growth Inv	7/04/95	US Equity Large Cap Growth	11,890,371	2.4
73	US1196281059	Buffalo Large Cap	19/05/95	US Equity Large Cap Growth	33,207,290	0.97
74	US9220383023	Vanguard Capital Opportunity Inv	14/08/95	US Equity Large Cap Growth	12,744,852,169	0.48
75	US5771191005	Matthew 25	16/10/95	US Equity Large Cap Growth	914,369,039	1.06
76	US5638216698	Manning & Napier Tax Managed A	1/11/95	US Equity Large Cap Growth	35,139,760	1.2
77	US74437E3053	Prudential Jennison Growth C	2/11/95	US Equity Large Cap Growth	2,706,761,210	1.76
78	US8523143019	STAAR Larger Company Stock	4/04/96	US Equity Large Cap Growth	3,421,983	1.94
79	US74316N1037	Profit	15/11/96	US Equity Large Cap Growth	17,396,588	1.35
80	US0079891067	American Trust Allegiance	11/03/97	US Equity Large Cap Growth	25,425,223	1.45
81	US36559B7082	Chesapeake Core Growth	29/09/97	US Equity Large Cap Growth	28,331,024	2.02
82	US19765H2489	Columbia Marsico Focused Equities C	31/12/97	US Equity Large Cap Growth	1,100,384,823	1.96
83	US19765H1986	Columbia Marsico Growth C	31/12/97	US Equity Large Cap Growth	2,051,918,162	1.93
84	US5730121010	Marsico Focus	31/12/97	US Equity Large Cap Growth	908,368,091	1.35
85	US5730122000	Marsico Growth	31/12/97	US Equity Large Cap Growth	578,300,603	1.37
86	US74440G3056	Prudential Jennison 20/20 Focus C	1/07/98	US Equity Large Cap Growth	2,258,569,820	1.88
87	US9300578724	Waddell & Reed Accumulative B	4/10/99	US Equity Large Cap Growth	1,304,893,211	2.47
88	US9300576587	Waddell & Reed Vanguard B	4/10/99	US Equity Large Cap Growth	1,441,400,281	2.5
89	US62826M4758	JPMorgan Growth Advantage B	29/10/99	US Equity Large Cap Growth	3,594,056,869	1.74
90	US01879K3095	AllianceBern Core Opportunities C	22/12/99	US Equity Large Cap Growth	153,139,370	1.98
91	US74440K7028	Prudential Jennison Select Growth C	1/06/00	US Equity Large Cap Growth	378,557,236	1.99
92	US0188661039	Iman K	30/06/00	US Equity Large Cap Growth	61,472,188	1.59
93	US4660006192	Ivy Large Cap Growth B	6/07/00	US Equity Large Cap Growth	1,481,296,248	2.06
94	US7469358409	Quaker Strategic Growth C	11/07/00	US Equity Large Cap Growth	169,909,972	2.99
95	US7799183097	T. Rowe Price Tax-Efficient Equity	29/12/00	US Equity Large Cap Growth	147,465,860	0.89
96	US36158T8861	GE Instl Premier Growth Equity Svc	3/01/01	US Equity Large Cap Growth	427,980,650	0.63
97	US45775L4086	T. Rowe Price Instl Large Cap Growth	31/10/01	US Equity Large Cap Growth	10,715,260,348	0.56
98	US7429355479	FundX Upgrader	1/11/01	US Equity Large Cap Growth	277,875,174	1.24
99	US4613741002	The Investment House Growth	28/12/01	US Equity Large Cap Growth	61,621,403	1.44
100	US3998747006	American Funds Growth Fund of Amer 529C	15/02/02	US Equity Large Cap Growth	143,050,293,211	1.56

#	ISIN	Name	Inception date	Global category	Fund size (in $)	Annual net expense ratio
101	US4099028148	JHancock US Global Leaders Growth C	20/05/02	US Equity Large Cap Growth	1,184,220,612	1.98
102	US77954Q3048	T. Rowe Price Blue Chip Gr R	30/09/02	US Equity Large Cap Growth	24,421,787,439	1.25
103	US4115118196	Harbor Capital Appreciation Inv	1/11/02	US Equity Large Cap Growth	24,159,615,030	1.02
104	US00078H4149	ASTON/Montag & Caldwell Growth R	31/12/02	US Equity Large Cap Growth	4,912,440,379	1.29
105	US85917L7597	Sterling Capital Special Opps Eq C	2/06/03	US Equity Large Cap Growth	1,042,862,310	1.99
106	US45775L5075	T. Rowe Price Instl Large Cap Core Gr	30/09/03	US Equity Large Cap Growth	1,187,420,889	0.6
107	US89154X4016	Touchstone Large Cap Growth C	6/10/03	US Equity Large Cap Growth	1,000,003,165	1.98
108	US92646A7081	Victory Large Cap Growth C	31/12/03	US Equity Large Cap Growth	204,875,504	2.1
109	US72387T6038	Pioneer Oak Ridge Large Cap Growth C	17/02/04	US Equity Large Cap Growth	91,186,648	2.1
110	US3786907705	Glenmede Large Cap Growth	27/02/04	US Equity Large Cap Growth	395,137,718	0.87
111	US92912M1053	Voya Corporate Leaders Trust B	18/11/35	US Equity Large Cap Value	1,701,830,210	0.5
112	US2562191062	Dodge & Cox Stock	4/01/65	US Equity Large Cap Value	58,448,842,168	0.52
113	US2174581080	Copley	1/09/78	US Equity Large Cap Value	71,116,761	1.6
114	US8085301096	Schwartz Value	30/06/83	US Equity Large Cap Value	32,435,987	1.45
115	US8360831056	Sound Shore Investor	17/05/85	US Equity Large Cap Value	2,483,057,955	0.93
116	US74155F1049	Primary Trend	15/09/86	US Equity Large Cap Value	17,863,626	1.99
117	US4377692011	Homestead Value	19/11/90	US Equity Large Cap Value	902,930,367	0.64
118	US9695572068	FBP Equity & Dividend Plus	30/07/93	US Equity Large Cap Value	29,421,930	1.07
119	US00142J3471	Invesco Growth and Income C	2/08/93	US Equity Large Cap Value	9,538,565,098	1.56
120	US00143M6874	Invesco Comstock C	26/10/93	US Equity Large Cap Value	13,079,625,357	1.61
121	US72366V2079	Pioneer Equity Income B	4/04/94	US Equity Large Cap Value	1,578,964,768	2.22
122	US76628R6568	RidgeWorth Large Cap Value Equity C	1/06/95	US Equity Large Cap Value	2,370,514,756	1.71
123	US57681T1025	Matrix Advisors Value	1/07/96	US Equity Large Cap Value	81,177,633	0.99
124	US0079898823	Edgar Lomax Value	12/12/97	US Equity Large Cap Value	52,305,258	0.76
125	US87234N4925	TCW Relative Value Large Cap N	2/01/98	US Equity Large Cap Value	744,481,395	1.13
126	US66538A3564	Al Frank Inv	2/01/98	US Equity Large Cap Value	97,111,865	1.49
127	US17800P7042	City National Rochdale Div & Inc N	1/06/99	US Equity Large Cap Value	189,404,980	1.16
128	US3048711069	Fairholme	29/12/99	US Equity Large Cap Value	8,676,640,196	1.01
129	US45775L2007	T. Rowe Price Instl Large Cap Value	31/03/00	US Equity Large Cap Value	2,189,826,989	0.58
130	US41012R8604	Hancock Horizon Value C	31/05/00	US Equity Large Cap Value	206,087,410	1.98
131	US93005P6097	Waddell & Reed Value B	15/12/00	US Equity Large Cap Value	871,455,723	2.54
132	US5529668551	MFS® Instl Large-Cap Value	1/05/01	US Equity Large Cap Value	184,914,488	0.65
133	US0250767615	American Century Value C	4/06/01	US Equity Large Cap Value	3,790,821,628	1.98
134	US0250767466	American Century Equity Income C	13/07/01	US Equity Large Cap Value	10,308,601,960	1.93
135	US88166L8357	TETON Westwood Income C	26/11/01	US Equity Large Cap Value	8,827,384	2.75
136	US1048261028	Queens Road Value	13/06/02	US Equity Large Cap Value	37,550,235	0.95
137	US55273H8824	MFS® Value 529C	31/07/02	US Equity Large Cap Value	34,487,561,216	1.72
138	US7795473062	T. Rowe Price Equity Income R	30/09/02	US Equity Large Cap Value	30,521,456,598	1.2
139	US87244W4666	TIAA-CREF Large-Cap Value Retail	1/10/02	US Equity Large Cap Value	5,269,312,915	0.8
140	US4099027645	JHancock Classic Value C	11/11/02	US Equity Large Cap Value	2,989,114,136	2
141	US92914G6504	VY T. Rowe Price Equity Income A	15/01/04	US Equity Large Cap Value	1,503,426,236	1.24
142	US92914G6686	VY Invesco Growth and Income A	20/02/04	US Equity Large Cap Value	652,517,509	1.23
143	US9204571080	Value Line Premier Growth	30/05/56	US Equity Mid Cap	392,647,805	1.24
144	US6537351008	Nicholas	14/07/69	US Equity Mid Cap	2,746,700,597	0.73
145	US8297961018	Sit Mid Cap Growth	2/09/82	US Equity Mid Cap	181,176,199	1.25
146	US9264643712	Victory Established Value R	16/08/83	US Equity Mid Cap	2,203,384,928	1.22
147	US0403371075	Ariel Investor	6/11/86	US Equity Mid Cap	2,139,318,591	1.03
148	US0682781002	Baron Asset Retail	12/06/87	US Equity Mid Cap	2,782,326,094	1.32
149	US7617241036	Reynolds Blue Chip Growth	10/08/88	US Equity Mid Cap	184,859,212	1.58
150	US5430692070	Longleaf Partners Small-Cap	21/02/89	US Equity Mid Cap	4,436,012,696	0.91

#	ISIN	Name	Inception date	Global category	Fund size (in $)	Annual net expense ratio
151	US06828M1080	Baron Partners Retail	31/01/92	US Equity Mid Cap	1,835,334,053	1.38
152	US7429351098	Hodges	9/10/92	US Equity Mid Cap	493,870,457	1.34
153	US00141M7965	Invesco Mid Cap Core Equity B	1/04/93	US Equity Mid Cap	2,625,617,768	1.91
154	US9204541059	Value Line Small Cap Opportunities	23/06/93	US Equity Mid Cap	356,886,636	1.26
155	US8888948473	Delafield Fund	19/11/93	US Equity Mid Cap	1,738,257,935	1.21
156	US7508692081	Rainier Small/Mid Cap Equity Original	10/05/94	US Equity Mid Cap	1,621,029,348	1.32
157	US82980D3017	Sit Small Cap Growth	1/07/94	US Equity Mid Cap	99,495,731	1.5
158	US1087471066	Bridgeway Aggressive Investors 1	5/08/94	US Equity Mid Cap	260,170,262	0.74
159	US89154X8728	Touchstone Mid Cap Growth C	3/10/94	US Equity Mid Cap	756,147,847	2.11
160	US90330L1052	US Global Investors Holmes Macro Trends	17/10/94	US Equity Mid Cap	51,994,593	1.96
161	US0682782091	Baron Growth Retail	30/12/94	US Equity Mid Cap	8,348,818,643	1.3
162	US45775L1017	T. Rowe Price Instl Mid-Cap Equity Gr	31/07/96	US Equity Mid Cap	4,315,899,429	0.61
163	US8855727355	IMS Capital Value	5/08/96	US Equity Mid Cap	40,466,319	2.06
164	US64122M6057	Neuberger Berman Genesis Adv	2/04/97	US Equity Mid Cap	14,656,483,435	1.38
165	US38142V6965	Goldman Sachs Mid Cap Value C	15/08/97	US Equity Mid Cap	10,383,518,420	1.89
166	US7599371058	Global IPO	19/12/97	US Equity Mid Cap	12,382,287	2.5
167	US9613231029	Westport R	31/12/97	US Equity Mid Cap	665,804,106	1.23
168	US9613232019	Westport Select Cap R	31/12/97	US Equity Mid Cap	389,543,526	1.37
169	US32008F8462	First Eagle Fund of America C	2/03/98	US Equity Mid Cap	3,161,378,743	2.16
170	US4976471072	Kirr Marbach Partners Value	31/12/98	US Equity Mid Cap	77,026,136	1.45
171	US4311137032	HighMark Geneva Mid Cap Growth B	4/01/99	US Equity Mid Cap	1,439,548,770	1.98
172	US9300576173	Waddell & Reed New Concepts B	4/10/99	US Equity Mid Cap	1,805,943,690	2.62
173	US66538E1010	NLFT III Lifetime Achievement	5/07/00	US Equity Mid Cap	207,468,093	1.19
174	US4660005699	Ivy Mid Cap Growth B	6/07/00	US Equity Mid Cap	4,749,585,604	2.1
175	US9499154253	Wells Fargo Advantage Common Stock C	30/11/00	US Equity Mid Cap	1,723,371,069	2.01
176	US3163898733	Fidelity® Leveraged Company Stock	19/12/00	US Equity Mid Cap	5,474,681,785	0.82
177	US3158053906	Fidelity Advisor® Leveraged Co Stk B	27/12/00	US Equity Mid Cap	4,795,463,090	1.91
178	US44134R8759	Hotchkis & Wiley Mid-Cap Value C	2/01/01	US Equity Mid Cap	3,566,914,138	2.07
179	US7468022634	Putnam Multi-Cap Value C	16/01/01	US Equity Mid Cap	420,075,144	1.86
180	US29372K5911	Gabelli Entpr Mergers & Acquisitions C	28/02/01	US Equity Mid Cap	269,134,529	2.24
181	US3551488758	Franklin Balance Sheet Investment B	1/03/01	US Equity Mid Cap		1.74
182	US1195301031	Buffalo Discovery	16/04/01	US Equity Mid Cap	649,584,331	1.01
183	US3141726519	Federated Kaufmann C	24/04/01	US Equity Mid Cap	5,649,376,345	2.5
184	US3391285067	JPMorgan Mid Cap Value C	30/04/01	US Equity Mid Cap	15,684,435,610	1.74
185	US4165285119	Hartford MidCap Value HLS IB	30/04/01	US Equity Mid Cap	504,639,534	1.09
186	US4166462063	Hartford MidCap Value B	30/04/01	US Equity Mid Cap	501,776,726	2.1
187	US8085302086	Ave Maria Catholic Values	1/05/01	US Equity Mid Cap	255,658,045	1.42
188	US3202698557	First Focus Growth Opportunities RetB	31/07/01	US Equity Mid Cap		2.05
189	US76628R5818	RidgeWorth Mid-Cap Value Equity C	30/11/01	US Equity Mid Cap	4,163,676,843	1.76
190	US1195302021	Buffalo Mid Cap	17/12/01	US Equity Mid Cap	607,284,090	1.01
191	US92914K6441	VY JPMorgan Mid Cap Value A	1/05/02	US Equity Mid Cap	739,969,231	1.36
192	US92914K6102	VY Baron Growth A	1/05/02	US Equity Mid Cap	1,114,856,051	1.49
193	US0155657242	Alger SMid Cap Growth B	8/05/02	US Equity Mid Cap	1,139,270,642	2.07
194	US3547134714	Franklin Small-Mid Cap Growth B	1/07/02	US Equity Mid Cap		1.74
195	US77957Y2054	T. Rowe Price Mid-Cap Value R	30/09/02	US Equity Mid Cap	12,187,066,306	1.31
196	US7795563078	T. Rowe Price Mid-Cap Growth R	30/09/02	US Equity Mid Cap	23,591,326,087	1.29
197	US87244W8477	TIAA-CREF Mid-Cap Value Retail	1/10/02	US Equity Mid Cap	5,174,286,760	0.76
198	US8855727272	IMS Dividend Growth	5/11/02	US Equity Mid Cap	8,750,600	1.97
199	US7809055509	Royce Premier Consult	2/06/03	US Equity Mid Cap	6,923,202,686	2.13
200	US3144654023	FAM Equity-Income Adv	1/07/03	US Equity Mid Cap		2.4

#	ISIN	Name	Inception date	Global category	Fund size (in $)	Annual net expense ratio
201	US9367936039	Wasatch Heritage Growth	18/06/04	US Equity Mid Cap	130,670,802	0.95
202	US94975J2353	Wells Fargo Advantage C&B Mid Cap Val C	26/07/04	US Equity Mid Cap	213,749,328	1.95
203	US74925K7063	Robeco WPG Small/Micro Cap Value	30/03/72	US Equity Small Cap	47,823,253	1.54
204	US9264643894	Victory Small Company Opportunity R	16/08/83	US Equity Small Cap	2,282,581,918	1.56
205	US0862331032	Berwyn	4/05/84	US Equity Small Cap	446,377,021	1.18
206	US4223591094	Heartland Value	28/12/84	US Equity Small Cap	1,291,884,161	1.08
207	US9367722017	Wasatch Core Growth	8/12/86	US Equity Small Cap	985,449,455	1.21
208	US9367721027	Wasatch Small Cap Growth	8/12/86	US Equity Small Cap	2,414,794,853	1.23
209	US6743751009	Oberweis Emerging Growth	7/01/87	US Equity Small Cap	58,845,022	1.53
210	US00170K2087	AMG Managers Skyline Special Equities	23/04/87	US Equity Small Cap	1,404,396,176	1.33
211	US00758M2200	ICM Small Company	20/04/89	US Equity Small Cap	1,283,986,237	0.93
212	US01877E2063	AllianceBern Small Cap Growth B	17/09/90	US Equity Small Cap	1,643,478,194	2.12
213	US3786906061	Glenmede Small Cap Equity Adv	1/03/91	US Equity Small Cap	1,520,863,747	0.91
214	US1152918330	Brown Capital Mgmt Small Co Inv	23/07/92	US Equity Small Cap	2,739,274,848	1.25
215	US70472Q4010	Pear Tree Columbia Small Cap Ord	3/08/92	US Equity Small Cap	111,572,074	1.53
216	US8631371056	Stratton Small Cap Value	12/04/93	US Equity Small Cap	1,393,619,967	1.15
217	US5018858002	LKCM Aquinas Small Cap	3/01/94	US Equity Small Cap	13,984,214	1.5
218	US6651624008	Northern Small Cap Value	31/03/94	US Equity Small Cap	2,722,219,386	1
219	US00758M2127	Rice Hall James Micro Cap Instl	1/07/94	US Equity Small Cap	41,845,185	1.51
220	US1087473047	Bridgeway Ultra-Small Company	5/08/94	US Equity Small Cap	151,742,214	1.17
221	US9367725085	Wasatch Micro Cap	19/06/95	US Equity Small Cap	328,677,509	2.13
222	US3551482066	Franklin MicroCap Value A	12/12/95	US Equity Small Cap	552,209,111	1.15
223	US6743752098	Oberweis Micro-Cap	29/12/95	US Equity Small Cap	54,782,015	1.86
224	US8523144009	STAAR Smaller Company Stock	4/04/96	US Equity Small Cap	3,983,459	1.94
225	US00758M1962	Rice Hall James Small Cap Instl	1/11/96	US Equity Small Cap	78,364,996	1.41
226	US0930016003	William Blair Small Cap Value N	23/12/96	US Equity Small Cap	600,881,188	1.49
227	US0189187307	AllianzGI NFJ Small-Cap Value C	20/01/97	US Equity Small Cap	7,922,062,320	1.93
228	US4834382065	Kalmar Growth-with-Value Sm Cp Inv	11/04/97	US Equity Small Cap	842,415,106	1.29
229	US76628R4589	RidgeWorth Small Cap Value Equity C	6/06/97	US Equity Small Cap	1,692,143,855	1.87
230	US7809058164	Royce Pennsylvania Mutual Consult	18/06/97	US Equity Small Cap	6,816,686,058	1.94
231	US0155658232	Alger Small Cap Growth C	1/08/97	US Equity Small Cap	257,283,753	2.16
232	US04314H5019	Artisan Small Cap Value Investor	29/09/97	US Equity Small Cap	2,364,439,544	1.24
233	US0682783081	Baron Small Cap Retail	30/09/97	US Equity Small Cap	5,720,488,681	1.31
234	US9367932079	Wasatch Small Cap Value	17/12/97	US Equity Small Cap	306,722,218	1.26
235	US4377695089	Homestead Small Company Stock	4/03/98	US Equity Small Cap	987,138,080	0.91
236	US6943368018	Pacific Advisors Small Cap Value C	3/04/98	US Equity Small Cap	224,034,621	3.08
237	US7429357111	Perkins Discovery	9/04/98	US Equity Small Cap	12,215,717	2
238	US1198041022	Buffalo Small Cap	14/04/98	US Equity Small Cap	3,597,336,490	1
239	US7492553372	Robeco Boston Partners Sm Cap Val II Inv	30/06/98	US Equity Small Cap	215,609,632	1.54
240	US55302C1027	MH Elite Small Cap Fund of Funds	1/09/98	US Equity Small Cap	6,424,885	1.25
241	US7492553943	Schneider Small Cap Value	2/09/98	US Equity Small Cap	63,613,631	1.15
242	US4702595087	James Small Cap	2/10/98	US Equity Small Cap	163,703,314	1.5
243	US7809057661	Royce Select I Invmt	18/11/98	US Equity Small Cap	47,918,655	1.2
244	US9219434035	Vanguard Tax-Managed Small Cap Adm	25/03/99	US Equity Small Cap	3,560,907,585	0.12
245	US7492551707	Bogle Small Cap Growth Inv	1/10/99	US Equity Small Cap	248,779,357	1.35
246	US0930014776	William Blair Small Cap Growth N	27/12/99	US Equity Small Cap	545,400,774	1.5
247	US26203E8518	Dreyfus/The Boston Co Small Cap Val I	1/02/00	US Equity Small Cap	352,540,989	0.99
248	US15649P1093	Century Small Cap Select Inv	24/02/00	US Equity Small Cap	419,496,281	1.41
249	US45775L3096	T. Rowe Price Instl Small-Cap Stock	31/03/00	US Equity Small Cap	1,545,612,044	0.68
250	US7795724035	T. Rowe Price Small-Cap Stock Adv	31/03/00	US Equity Small Cap	9,935,524,670	1.19

#	ISIN	Name	Inception date	Global category	Fund size (in $)	Annual net expense ratio
251	US77957Q2021	T. Rowe Price Small-Cap Value Adv	31/03/00	US Equity Small Cap	9,873,025,281	1.09
252	US3547136461	Franklin Small Cap Growth C	1/05/00	US Equity Small Cap	1,711,306,404	1.9
253	US3176092795	Emerald Growth C	30/06/00	US Equity Small Cap	257,648,255	1.94
254	US7469357419	Quaker Small-Cap Value C	28/07/00	US Equity Small Cap	38,253,807	2.68
255	US7809055921	Royce Total Return Consult	16/10/01	US Equity Small Cap	5,217,805,781	2.15
256	US7809056184	Royce Heritage Consult	7/12/01	US Equity Small Cap	378,051,945	2.37
257	US41012R8117	Hancock Horizon Burkenroad Small Cap D	31/12/01	US Equity Small Cap	851,588,608	1.65
258	US94975P8766	Wells Fargo Advantage Small Co Value B	31/01/02	US Equity Small Cap	113,087,622	2.2
259	US19765J7726	Columbia Small Cap Value Fund II C	30/04/02	US Equity Small Cap	1,860,868,528	2.04
260	US1048262018	Queens Road Small Cap Value	13/06/02	US Equity Small Cap	77,751,178	1.24
261	US4115116943	Harbor Small Cap Value Investor	1/11/02	US Equity Small Cap	585,754,154	1.21
262	US0155707088	Alger Small Cap Growth Institutional R	27/01/03	US Equity Small Cap	850,737,787	1.73
263	US0858911094	Pinnacle Value	1/04/03	US Equity Small Cap	65,896,762	1.46
264	US7809055434	Royce Special Equity Consult	2/06/03	US Equity Small Cap	3,351,335,528	2.17
265	US5018856022	LKCM Small Cap Equity Advisor	5/06/03	US Equity Small Cap	1,033,202,145	1.2
266	US9367935049	Wasatch Micro Cap Value	28/07/03	US Equity Small Cap	186,048,636	2.25
267	US00170K5544	AMG SouthernSun Small Cap Investor	1/10/03	US Equity Small Cap	901,989,794	1.22
268	US03332V8072	Ancora Special Opportunity C	5/01/04	US Equity Small Cap	11,334,856	2.59
269	US94975J3344	Wells Fargo Advantage Small Co Growth C	30/01/04	US Equity Small Cap	362,095,485	2.2
270	US72387T2078	Pioneer Oak Ridge Small Cap Growth B	17/02/04	US Equity Small Cap	2,225,796,948	2.3
271	US3499032296	Adams Harkness Small Cap Growth	27/02/04	US Equity Small Cap	26,796,911	1.8
272	US9045045377	Undiscovered Mgrs Behavioral Value B	4/06/04	US Equity Small Cap	1,200,799,657	1.95

Part II

Social responsibility

Chapter 4

Multi-criteria decision analysis: Methods to define and evaluate socially responsible investments

Abstract

Originally being a niche strategy followed by few investors, socially responsible investing (SRI) now represents a significant part of the assets under management. After summarizing empirical evidence on the performance of SRI funds, we present four challenges that are facing the further development of SRI and point to multi-criteria decision analysis (MCDA) as the methodological framework that could help overcome these challenges. A first group of challenges calls for the development of a social performance indicator, which can score and classify mutual funds with respect to social responsibility. Another challenge requires a transparent tool for retail investors interested in SRI to learn about their SRI preferences. Reviewing the three schools of available MCDA methods, we present a concrete approach for future research in building such a social performance indicator and a retail investor tool for SRI.

4.1 Introduction

Socially responsible investing (SRI) has experienced a rapid growth over the past decade, reflecting the increasing awareness of investors for environmental, social and governance (ESG) issues. Different definitions of SRI have been suggested. Instead of looking at one static definition, we might consider the development of the SRI concept. The first generation of SRI simply consisted of the application of negative screens to the investment universe. In considering different investment opportunities, some criteria are established to screen out "sin stocks" or companies that are in discordance with a set of moral and/or ethical principles. In the second generation of SRI, the focus was more on adopting positive screens and a best-in-class approach. The combination of both positive and negative screens led to the third generation of SRI. The fourth and most recent generation of SRI includes shareholder activism, next to the application of positive and negative screens.

The origins of SRI go back to the moral principles adopted by religious organizations in considering investment alternatives. The definitive breakthrough for socially responsible investing came with the massive worldwide protest against the racist system of apartheid in South Africa. In recent years SRI has moved from niche to mainstream (KPMG & ALFI, 2013), as issues like global warming, the Kyoto Protocol, corporate governance, and community investing have gained significant attention from investors around the world. In addition, governments in western countries have taken many regulatory initiatives to stimulate SRI. Both elements create a pro-SRI environment in which SRI will continue to grow and establish its relative importance as an asset class (Renneboog, Ter Horst, & Zhang, 2008a). According to the Forum for Sustainable and Responsible Investment (2012) a little more than one out of every nine dollars (11.3%) under professional asset management in the United States is invested in the SRI universe. At the start of 2012, SRI assets managed by professionals stood at \$3.744 trillion, a rise of more than 486 percent from the \$639 billion in 1995. Over the same period, the broader universe of assets under conventional professional management rose only 376 percent. The latest SRI study by Eurosif (2012) demonstrates similar results. The combined growth of SRI strategies has outperformed the conventional market in Europe, despite the financial turmoil from the recent past.

The central aim of this paper is to open up alleys for future research, by pointing out some outstanding challenges in the SRI field and by presenting a methodological approach to address these challenges. We briefly summarize the empirical research on the financial performance of SRI funds and argue that the current dichotomic classification of a funds' social responsibility explains why previous research finds insignificant differences in returns between SRI and conventional funds. In order to enrich the academic state-of-the-art, we see a need for a proper framework to select, evaluate and categorize SRI funds in a more nuanced and continuous way. Such a framework can also benefit governmental agencies in regulating the SRI market and commercial banks in developing new SRI products. Additionally, from the 2012 Eurosif SRI study, we see that the European SRI retail market is particularly underdeveloped. The main part of SRI investments and growth come from institutional investors. One of the reasons is a lack of transparency for retail investors and the proliferation of different methods to determine the social performance of mutual funds. Clearly, there is a need for a transparent yardstick allowing retail investors to determine the social responsibility content of their investments. Also, retail investors need some guidance in getting to know their personal preferences with regard to social investing. For all of these challenges, we point to multi-criteria decision analysis (MCDA) as an interesting methodology to help define and evaluate the social performance of mutual funds. We review the different schools of MCDA methods and present a concrete approach to apply MCDA to the modern challenges facing the further development of SRI as an asset class.

4.2 SRI performance debate

As several review papers on SRI performance have been written (e.g. Margolis & Walsh, 2003; Orlitzky, Schmidt, & Rynes, 2003), it is not our aim to provide a complete overview of earlier research on this topic. Rather, we present a summary from the two fields where research has been conducted and the different findings that have been presented.

In strategic management science, the debate on corporate social versus corporate financial performance goes back to the opposing views of Friedman (1970) and Freeman (1984). In a New York Times Magazine article, Friedman (1970) makes the case for shareholder theory, which states that the sole responsibility of businesses is to maximize the value for its shareholders. In this view, which is also referred to as Friedman's doctrine, it is believed that society at large benefits most if companies simply focus on maximizing their own profits. Consequently, any corporate social responsibility (CSR) initiative is obsolete. Stakeholder theory, first proposed by Freeman (1984), takes a different view on the role of a business. The responsibility of a firm should not be limited to the shareholders, but should consider all the stakeholders. This is not only believed to increase overall welfare, but also the profitability of individual firms as the theory argues that a good relationship with all the stakeholders will improve long-term financial performance. Reconciliating the opposing views, Mackey, Mackey and Barney (2007) find that demand and supply conditions for socially responsible investment opportunities determine whether socially responsible decisions can lead to better financial performance.

The same question on the financial performance of SRI was posed in the field of financial economics. Mostly, researchers have implemented empirical models to compare the financial performance of SRI and conventional funds, controlling for different factors of risk. The first empirical studies go back to Moskowitz (1972) and Bragdon and Marlin (1972), who find a positive rank correlation between corporate social and financial performance.

The sophistication of the applied methods has since then increased, and so has the quality of the results. Following the rank correlation tests, SRI performance was researched using the capital asset pricing model (CAPM; Sharpe, 1964), controlling only for market risk (e.g. Hamilton, Jo, & Statman, 1993), and using performance ratios like the Sharpe index (e.g. Sauer, 1997). As more recent advances in empirical asset pricing got adopted across the field, the CAPM was gradually replaced by the Fama-French (1993) three-factor model and the Carhart (1997) four-factor model (e.g. Bauer, Koedijk, & Otten, 2005). Today, most SRI performance research implements a conditional four-factor model, following Ferson and Schadt (1996), taking into account possible time-varying risk. Given these different possible approaches, Geczy, Stambaugh and Levin (2003) found that the cost of investing in SRI funds instead of conventional funds crucially depends on two elements: the belief held by the investor regarding the valid underlying asset pricing model, and the ability of fund managers to select stocks. The cost of investing in SRI when assuming the CAPM holds true and fund managers have no stock-picking skills is negligible. However, the cost is substantial when adhering to a four-factor model and assuming that fund managers have some skills in selecting stocks.

Regardless of the evolution in the methodological approach of researching the performance of SRI funds, inconclusive results have been found from the start. This led scholars to divide into believers and non-believers of SRI, each with their own set of arguments. The non-believers, who argue that SRI funds can only underperform traditional funds, mainly refer to modern portfolio theory (Markowitz, 1952). Because social screening adds constraints to the optimization problem of finding an efficient portfolio, some idiosyncratic risk cannot be diversified away. Consequently, SRI portfolios are not on the efficient frontier and thus will yield subpar risk-adjusted returns. The critique of the non-believers is also in line with the argument of Friedman (1970). Engaging in socially responsible activities increases the operational costs of a firm, which negatively impacts overall profitability. The believers contend that social responsibility is not a cost, but rather an investment as firms are an inherent part of their social environment (Granovetter, 1985). This view is supported by stakeholder theory (Freeman, 1984). Additionally, the believers refute the argument that social screening leads to inefficient portfolios. Even though the pool of potential stocks to diversify away idiosyncratic risk is smaller because of the additional constraints, the quality of this pool is believed to be higher as the screening process yields value-relevant information for the investor (Barnett & Salomon, 2006).

4.3 SRI challenges and issues

Despite the progress of SRI as an asset class, the research on SRI performance is still ongoing. We identify four important challenges for the continued growth of SRI, both in the academic and the professional world.

A first challenge concerns the methodological approach for the performance analysis of SRI funds in academics. From the review of methodologies to test for SRI performance, we learn that measuring risk-adjusted returns from asset pricing models is standard practice today. To compare SRI and conventional funds, a difference portfolio is usually constructed from a dichotomous dummy variable that indicates whether the fund is labeled socially responsible or not. The problem with the dummy approach is that it neglects possible heterogeneity among different SRI funds and that it does not take into account the multiple dimensions relevant to social responsibility. In reality socially responsible investors do not adopt a dichotomous classification approach and need to carefully examine the mutual funds' prospectus to examine if the fund's investment strategy and social responsible guidelines meet their individual ethical standards (Hollingworth, 1998). Hoggett and Nahan (2002) and Tippet (2001) show that this kind of SRI investment information might be hard to retrieve or even be unreliable. Barnett and Salomon (2006) already addressed the issue of dichotomy and found a curvilinear relationship between corporate social and financial performance. These findings even suggest that both views on SRI performance could coexist. Schwartz (2003) and Koellner, Weber, Fenchel and Scholz (2005) propose a general code of ethics for socially responsible investing regarding the information disclosure or transparency, the investment process and the credibility of information, but there is a lack of specific social responsible indicators that capture the multifaceted nature of SRI. A methodology to overcome the dichotomous measurement of social responsibility in asset pricing models is still lacking, although some earlier research was devoted to examining the effect of different types of SRI screens on flow-return relations (Renneboog, Ter Horst, & Zhang, 2011). We believe that the development of a social responsibility indicator to score and/or classify funds based on multi-criteria decision analysis (MCDA) methodologies could overcome this issue. The MCDA field is devoted to the development of appropriate methodologies that can be used to support and aid decision makers in circumstances where multiple conflicting decision factors (objectives, goals, criteria) have to be considered simultaneously. The application of MCDA to finance problems is not new (see e.g. Steuer & Na (2003) and Zopounidis & Doumpos (2002) for an overview of this line of research).

The lack of a proper regulatory framework in the certification of mutual funds is the second challenge. Although at the company level, several independent agencies[1] try to supply transparent and credible information about the social, labor and environmental performance of corporations throughout the world, few rating agencies monitor the process-oriented social responsibility value/authenticity of mutual funds. Most of these agencies only provide financial information about the funds (costs, performance, risk and liquidity) and conventional investment strategy information (type of security, country and industry allocation, financial investment objectives and fund composition). Supervising authorities are currently unable to adequately screen the design of ethical mutual funds to, for example,

[1] Some examples are KLD, Ethibel, Vigeo, Innovest, Oekom Research, SAM, Sustainalytics, Corporate Monitoring, EthicScan Canada and EIRIS.

grant a certificate of "ethical authenticity" to funds or to promote ethical investments all together. As SRI is becoming more popular, this gives an incentive to investment institutions to label their mutual funds socially responsible, even though this is not really the case. The MCDA framework can again help to create a tool to classify mutual funds based on an assessment of a wide variety of underlying socially responsible criteria.

In the absence of a clear framework to define and categorize SRI funds, commercial banks also need to spend considerable time and resources to develop, implement and communicate an in-house SRI view. This forms the third challenge to the further development of SRI. A framework based on MCDA can help companies to construct SRI mutual funds in a more efficient, consistent and transparent way. Consequently, consumers would be less confused by standalone SRI definitions that differ from bank to bank and would be enabled to compare different SRI mutual funds in a straightforward manner. Assessing an investment alternative would then be possible along three dimensions: risk, return and social responsibility.

These first three challenges are rather similar, in the sense that they can be addressed in the same way. They all need an overarching framework that can help in scoring and classifying mutual funds based on social responsibility. More specifically, a social performance indicator can be helpful in discriminating between the social responsible design of mutual funds in a more continuous way, it would give regulators a tool to develop labels — based on categories or scores — for genuine SRI funds and it would facilitate the process of developing new SRI mutual funds for banks. This performance indicator needs to be as general as possible, taking into account views from all the different stakeholders and interest groups involved in the SRI field. A more concrete approach for developing this indicator is presented in Section 4.5.

A final outstanding challenge for the field of SRI was revealed by the 2012 Eurosif report on the European SRI mutual fund industry. Even though the SRI market continues to grow, the retail segment remains underdeveloped as growth and volume in the SRI market predominantly comes from institutional investors. For the overall European mutual fund industry, 25% of assets under management are held by retail investors (European Fund and Asset Management Association, 2013). Retail investments in the European SRI mutual fund industry only amount to 6% of assets under management, which is illustrative of a large potential for growth (Eurosif, 2012). As a main reason for this underdevelopment of the European SRI retail mutual fund market, Eurosif (2012) points towards bad communication and a lack of transparency and clarification of SRI strategies, which keeps many retail investors from investing in SRI funds. Again, MCDA could provide the framework to overcome this challenge, as the methodology can assist retail investors in handling extensive information in a transparent way. As MCDA is focused on accommodating better decisions, it can also assist retail investors in making wiser investment choices. Similar to the investment services directive by the European Commission to allow investors to better understand the risk they want to take, we believe the MCDA framework could be formalized into a "green" MiFID (Markets in Financial Instruments Directive) questionnaire, which could assist investors in better understanding their social responsibility preferences and increase the investor protection in Europe with respect to SRI (Davies, Dufour, & Scott-Quinn, 2006). Note that the MCDA tool to address this challenge needs to be more tailored to the needs of individual investors, which sets the fourth challenge apart from the first three.

Different authors have attempted to define and evaluate SRI before. Within the field of operations research, Pérez-Gladish and M'Zali (2010) already constructed a first SRI indicator applying one MCDA method called the analytic hierarchy process (AHP). Although very interesting, their paper lacks methodological sophistication as it only considers and implements one MCDA method, whereas other methods might be more appropriate. Additionally, the input required from investors and/or experts is too detailed (e.g. different gradations in assessing SRI criteria), which makes the method harder to implement, more unreliable and susceptible to rank reversal issues, which we discuss in more detail further on in this paper. Within the field of financial economics, different authors have tried to include social responsibility in performance evaluation estimations (e.g. Renneboog, Ter Horst, & Zhang, 2008b). However, these attempts are rather limited in scope, as they typically reduce the concept of social responsibility to a dummy-variable, which is driven by the underlying decision of a financial institution to promote investment products as SRI or not, leading to a classification bias. Finally, the above-mentioned SRI challenges have also been extensively addressed in the strategic management literature (e.g. Waddock & Graves, 1997). Although insightful, these studies fail to incorporate the social responsibility topic in the financial performance methodology, which leads to weak and incomplete results on the performance of socially responsible investment strategies vis-à-vis conventional investment strategies. CSR evaluation techniques are also not appropriate to address the problem at hand since they focus on in-depth individual analysis of companies. As we are trying to address SRI performance and regulatory issues on the fund level, we are looking for aggregate fund-level assessments of social responsibility, which calls for a different approach and set of criteria. Furthermore, we aim for a quantitative instrument transforming qualitative criteria into an indicator score or category, as to integrate these into performance evaluation multi-factor models from the financial economics literature.

Clearly, MCDA could be instrumental in addressing all of the above challenges, and in improving earlier work on this topic. This is not to say that MCDA is the magic formula that will resolve each and every issue, but merely that it can provide the framework to help move the SRI field to the next level. In order to further explore and guide future research, we review the different schools of MCDA methodologies and show how these methods could be used as a tool to define and evaluate SRI funds.

4.4 Overview of MCDA methods

The development of MCDA, an advanced field of operations research, is based on the simple finding that a single objective, goal, criterion or point of view is rarely used to make real-world decisions. The MCDA field is devoted to the development of appropriate methodologies that can be used to support and aid decision makers in situations where multiple conflicting decision factors (objectives, goals, criteria) have to be considered simultaneously. Given the different dimensions to the concept of social responsibility, MCDA is also relevant in scoring and classifying SRI mutual funds.

Within a multi-criteria context, decision-making problems are realized in the following paradigm: a decision maker considers a set of alternatives and seeks to take an "optimal" decision considering all the factors that are relevant to the analysis. Since these factors usually lead to conflicting results and conclusions, the "optimal" decision is not really optimal in the traditional optimization perspective. Instead, it is a satisfactory non-dominated decision, i.e. a decision that is in accordance with the decision maker's system of values and is not dominated by other possible decisions.

Irrespective of whether the set of alternatives A is discrete or continuous, making a decision in a multi-criteria context requires the appropriate aggregation of all the pertinent decision factors, which are referred to as "evaluation criteria" or simply "criteria". Formally, a criterion is a non-decreasing real-valued function that describes an aspect of the global performance of the alternatives and defines how the alternatives are compared to each other.

In making a decision within the multi-criteria context the aggregation of the criteria is a crucial process. This aggregation can be performed in many different ways depending on the form of the criteria aggregation model. Within the MCDA field one can distinguish three main forms of aggregation models: outranking relations (relational form), utility functions (functional form) and decision rules (symbolic form). The construction of an aggregation model is mainly of interest in the case where A is discrete. In such a case the alternatives are clearly identifiable and consequently their performance on each criterion can be specified rather easily. In the case where A is continuous, however, this is not a straightforward process, simply because it is impossible to identify all the alternatives that are relevant to the analysis. In this case special interactive aggregation techniques have been developed in MCDA to allow for the efficient search of the solution space.

In all cases, the aggregation of the criteria is performed so as to respect the decision maker's (DM) judgment policy. To ensure that this objective is achieved some information on the preferential system of the DM must be specified, such as the criteria weights. The required preferential information can be specified either through direct procedures in which a decision analyst elicits it directly from the DM, or through indirect procedures in which the DM provides examples of the decision situations that he/she faces and the decision analyst examines them to determine the required preferential parameters which are most consistent with the DM's global evaluations. The latter approach is known in the MCDA field as "preference disaggregation analysis" (Jacquet-Lagrèze & Siskos, 1982, 1983, 2001).

It is recognized that the MCDA models can be classified into three broad categories, or schools of thought: (1) value measurement models in which one decision option may be preferred to another and for which scores are developed initially for each individual criterion, and then synthesized in order to effect aggregation into higher level preference models; (2) outranking models in which alternative courses of action are compared pairwise, initially in terms of each criterion, in order to identify the extent to which a preference for one over the other can be asserted. In aggregating such preference information across all relevant criteria, the model seeks to establish the strength of evidence favoring selection of one alternative over the other; (3) goal, aspiration or reference level models in which desirable or satisfactory levels of achievement are established for each of the criteria. The process then seeks to discover options, which are in some sense closest to achieving these desirable goals or aspirations. Note that softer methods outside of these three categories exist as well. For example, the even-swap method by Hammond, Keeney and Raiffa (2000), which provides a very practical and understandable way of dealing with value tradeoffs to make a decision.

A number of authors have highlighted the similarities of data envelopment analysis (DEA) and MCDA models, commenting principally from a theoretical perspective on the mathematical structure and methods for solution. Given these similarities it is possible that the two approaches could be viewed as competing. DEA could be described as an approach, which seeks to extract as much as possible from “objective” historical data, without resort to subjectivity. In contrast, MCDA actively seeks to elicit, understand and manage value judgments.

4.4.1 Value measurement methods

The idea behind value measurement methods is to formulate a quantitative score for every alternative based on an aggregate value judgment of the relevant criteria (see e.g. Belton & Stewart, 2002). This score could then be used to rank or classify alternatives. The most straightforward approach is to score each alternative on every individual criterion, and then calculate a weighted sum of these partial scores based on the DM’s judgment of the relevant importance of each criterion.

Multi-attribute value theory

Multi-attribute value theory (MAVT) is an extension of basic value measurement methods that takes into account possible non-linearity of the preference functions. In a first step, a hierarchical value tree, which represents the hierarchy of relevant criteria in scoring different alternatives, has to be constructed. In constructing the value tree, it is important to consider the condition of preferential independence, which means that tradeoffs between different criteria should not depend on any other criteria. Once the alternatives have been determined, the second step consists of constructing a performance table by scoring the different alternatives with respect to the different criteria. If the decision maker feels comfortable with the alternatives and the criteria, the scoring process can be completed by direct assessment. If the decision maker has more difficulties with the scoring process, other scoring methods can be considered, e.g. indirect assessment, using qualitative scales or by pairwise comparison.

The distinctive feature of MAVT is in the elicitation of partial value functions for each criterion, which represent the utility derived by the DM from the performance of an alternative with respect to a single criterion. Deriving the true underlying value function is not straightforward. To help the DM in this complex process, value functions can be derived in an indirect way, for example, via standard differences or via bisection methods.

Next to determining the underlying partial value functions, weights of criteria need to be elicited. Again, different methodologies can be applied. Either the DM feels comfortable assessing the importance of criteria directly, or methods like preference disaggregation analysis are used (Jacquet-Lagrèze & Siskos, 1982, 1983, 2001). In a final step, the overall score for each alternative is calculated aggregating the partial value functions on the basis of the elicited weights for every criterion. Different forms of aggregation are possible (e.g. additive or multiplicative). Note that this final step is not the same as a simple weighted sum, since only the transformed utility values from the partial value functions are used for each criterion, and not the direct scores from the performance table itself.

Similar to MAVT is multi-attribute utility theory (MAUT), which is based on expected utility theory and calls for even stronger assumptions. The main advantage of MAUT is the possibility to take into account uncertainty and risk. However, this makes it even more complex to elicit final scores for the alternatives. Therefore, MAVT is still the preferred methodology in approaching real-life decision problems. For a more complete overview of these techniques, see for example Keeney and Raiffa (1993) and Wallenius et al. (2008) for some more recent accomplishments and future applications.

MAVT/MAUT is of particular interest to our decision problem because of the straightforward interpretation. In essence it can be considered a sophistication of the weighted sum approach using value/utility information, which is easy to understand for every investor. However, future research should consider the trade-off between the ease of understanding and the complexity of the underlying utility functions. Because the input for the model comes from a decision maker, i.e. the investor, simpler utility functions are always preferred to more complex functions to ensure practical implementation. A disadvantage of this method is the lack of a built-in methodology to determine criteria weights. No special software is required for implementation, as MAVT/MAUT can easily be programmed into a spreadsheet package.

Analytic hierarchy process

The analytic hierarchy process (AHP), presented in detail by Saaty (1980), is an elegant approach in its simplicity, for addressing and analyzing discrete alternative problems with multiple conflicting criteria. Like MAVT, the AHP starts by subdividing a problem into a hierarchy of overall objective criteria. As we work to build this AHP hierarchy, we increase our understanding of the problem as a whole. Particular about AHP is the use of pairwise comparisons to elicit the criteria weights from experts. Psychologists argue that it is easier and more accurate to express one's opinion on only two alternatives than simultaneously on all the alternatives. It also allows consistency cross checking between the different pairwise comparisons. Starting at the bottom level of the hierarchy, we conduct pairwise comparisons between the elements immediately below each other. Under real conditions, it is not difficult, based on the condition of transitivity, to identify improperly filled in questionnaires. The AHP method assesses the consistency of each expert opinion and defines a consistency index (Saaty, 1980, 2005).

One of AHP's strengths is the possibility to evaluate quantitative as well as qualitative criteria and alternatives on the same preference scale. These can be numerical, verbal or graphical. The use of verbal responses is intuitively appealing, user-friendly and more common in our everyday lives than numbers. Nevertheless, it may also allow some ambiguity in non-trivial comparisons, which has been criticized (Donegan, Dodd, & McMaster, 1992). To derive priorities, the verbal comparisons must be converted into numerical ones. In Saaty's AHP the verbal statements are converted into integers from one to nine. Theoretically there is no reason to be restricted to these numbers and verbal gradation. Although the verbal gradation has been little investigated, several other numerical scales have been proposed. Clearly, the choice of the "best" scale is a very heated debate. Some scientists therefore argue that the choice of scale depends on the person and the decision problem (Harker & Vargas, 1987; Pöyhönen, Hamalainen, & Salo, 1997).

AHP still suffers from some theoretical disputes. Rank reversal is surely the most debated problem. This phenomenon is still not fully resolved and may never be because the aggregation of preferences transposed from scales of different units is not easily interpretable and even questionable according to Roy (1996). In this sense, the rank reversal problem is not specific to AHP, but to the normalization of scores. The assumption of preferential independence may also be a limitation of AHP (and other MCDA methods). The analytic network process (ANP), a generalization of AHP with feedbacks to adjust weights, may be a solution. However the decision maker must answer a much larger number of questions, which may be complex (Saaty & Takizawa, 1986). A simplified ANP, while still keeping its proprieties, would be beneficial for a wider adoption of the method (Ishizaka & Labib, 2011). The choice of a hierarchy and a judgment scale is also important and difficult. Problem structuring methods could help in the construction of AHP hierarchies, which is its less formalized aspect (Petkov & Mihova-Petkova, 1997; Petkov, Petkova, Andrew, & Nepal, 2007).

Several works can be found in the literature relating AHP with finance. Beyond improving the quality of the decisions, the AHP is shown as a useful tool to support the process of examining, justifying, negotiating, and communicating ethical decisions. Pérez-Gladish and M'Zali (2010) already showed how AHP can be implemented to address the problem at hand. Most valuable about the AHP method is the built-in procedure to determine criteria weights using pairwise comparisons. This procedure can be used to complement other methodologies as well. With simple dichotomous utility functions, the MAUT/MAVT approach becomes nested within the AHP method when using pairwise comparisons to determine criteria weights. A common critique on the AHP method is that it is rather simplistic and only useful for its pairwise comparison approach. The use of a method from another school of thought, together with the AHP weighting procedure must therefore be considered in future research as a potential MCDA approach.

Other value measurement methods

Besides MAVT and AHP, other value measurement models exist. Like AHP, MACBETH (Measuring Attractiveness by a Categorical Based Evaluation TecHnique; Bana e Costa & Vansnick, 1994) is based on pairwise comparisons to evaluate alternatives with respect to different criteria. The difference, however, is in the use of an interval scale instead of a ratio scale and the fact that the MACBETH method calls for a much greater number of pairwise comparisons to elicit criteria weights, which makes it less desirable for future implementation in practice. MACBETH is also harder to implement in practice as it can only be achieved through costly software, whereas methods like AHP can be implemented using free software. The linear programming algorithm behind MACBETH does not always yield consistent results and is more black box and harder to understand than most other value measurement methods. Inconsistencies are however easier to spot in the MACBETH method. Therefore, we suggest that future research looks into the application of both AHP and MACBETH, considering criteria like stability intervals to decide which one works best. To further accommodate the elicitation of utility functions from decision maker preferences, disaggregation methods like UTA (UTilités Additives; Jacquet-Lagreze & Siskos, 1982) and UTADIS (UTilités Additives DIScriminantes; Jacquet-Lagrèze & Siskos, 1982; Zopounidis & Doumpos, 1999) have also been developed, but might be harder to implement for the purpose of our decision problem.

4.4.2 Outranking methods

ELECTRE

The ELECTRE (ELimination Et Choix Traduisant la REalité) family of methods by Roy (1985) is based on the concept of outranking: “one solution outranks another if it is at least as good as the other in most respects, and not too much worse in any one respect." Typical for the outranking approach is the unicriterion comparison of alternatives based on preference degrees. Such preference degrees express how a decision maker prefers one alternative to another, based on an underlying preference function. A typical ELECTRE analysis yields only an outranking relation of the different alternatives, and no concrete quantitative output. Therefore, it might be slightly less appropriate to build a social performance indicator. Just like AHP, ELECTRE also suffers from the rank reversal issue (Wang & Triantaphyllou, 2008). However, ELECTRE has been successfully implemented in financial research before. For instance, Martel, Khoury and Bergeron (1988) employ ELECTRE to study the limitations of conventional risk in being able to capture global risk in a portfolio context. Also influenced by ELECTRE is BANK ADVISER by Mareschal and Brans (1991), which has been successful in the banking industry. Members of the ELECTRE family are ELECTRE I, II, III, IS, IV and TRI.

An important limitation of the ELECTRE methodology is that it does not generally yield a scoring output. Therefore, application of the ELECTRE method would call for some additional transformation to obtain the scores or categories we need for further implementation in performance regressions. Different software packages are available to implement this method. ELECTRE is also considered a more outdated technique in comparison to PROMETHEE, which should be taken into account when comparing methods upon implementation. The need for information from the decision maker can also be quite considerable in the ELECTRE methodology.

PROMETHEE

Originally developed by Brans and Vincke (1985), PROMETHEE (Preference Ranking Organization METHod for Enrichment Evaluations) is also based on the theory of outranking relations. The outranking methods include two phases: the construction of an outranking relation, and the exploitation of this relation in order to assist the decision-maker. To capture the outranking relation, Brans, Vincke and Mareschal (1986) use six types of functions that cover most of the cases occurring in practical applications. The basic principles of the PROMETHEE method in relation with other methods of the same field are the following: extension of the notion of criteria, a valued outranking relation and exploitation of the outranking relation. In the PROMETHEE method the valued outranking relation is less sensitive to small modifications and its interpretation is straightforward. The exploitation of the valued outranking relation of the PROMETHEE method refers to the case in which the alternatives have to be ranked from best to worst. A typical PROMETHEE analysis yields a quantitative output in the form of net flows (valued outranking relation), which represent the relative preference of the DM for one alternative to another. This quantitative output could be transformed to an index, which makes PROMETHEE a viable candidate to build a social performance indicator. Again, rank reversal occurs as an issue when implementing the PROMETHEE method (e.g. Mareschal, De Smet, & Nemery, 2008), which is inherent to the normalization of scores as mentioned before. As an example of the application of PROMETHEE in investment decision-making, we refer to Qu, Li and Pei (2012).

Implementation of PROMETHEE is possible using either commercial software packages (e.g. D-Sight) or free academic alternatives (e.g. Visual PROMETHEE). In contrast to the ELECTRE method, PROMETHEE yields a quantitative output in the form of net flows, which might need some transformation/scaling to get a desirable score/category for use in further financial analyses. The PROMETHEE analysis can also be used in conjunction with a graphical visualization procedure termed GAIA (Geometric Analysis for Interactive Aid), which provides a two-dimensional representation of the multi-dimensional problem (Brans & Mareschal, 1990). Such a visual aid is instrumental in presenting results in a compact way and in gaining more insights into how scores and rankings were calculated, and how performances of individual alternatives could be improved. A downside of PROMETHEE, just like ELECTRE and MAUT/MAVT, can be the greater need for input from the decision maker. Members of the PROMETHEE family are PROMETHEE I, II, III, IV, V, VI and GDSS.

4.4.3 Goal, aspiration or reference level methods

The last school of MCDA methods covers a wide range of optimization methods, most of which require highly quantifiable inputs. The advantage of goal programming (GP) and aspiration methods is that multiple criteria can be incorporated into a model that can be solved using conventional (single criterion) optimization software. The disadvantage is that information about the decision maker's preferences is required a priori in the form of priority levels, importance weights, and goal target values. GP models can be divided into linear goal programming models, interactive multiple goal programming (IMGP) and interactive sequential goal programming (ISGP). These kinds of models are especially useful when the set of alternatives is continuous. For example, GP might come in handy when determining the optimal funds-of-funds strategy or when managing a portfolio of socially responsible investments (e.g. Hallerbach, Ning, Soppe, & Spronk, 2004).

Next to the optimization methods that are also found in the broader field of operations research (OR), this third school of methods also includes reference level models like TOPSIS (Technique of Order Preference Similarity to the Ideal Solution). The central principle of TOPSIS is very simple, and also allows for the use of qualitative criteria, as long as they can be translated onto a numeric scale (Behzadian, Otaghsara, Yazdani, & Ignatius, 2012). With TOPSIS, the distance from every considered alternative to a theoretically defined worst and ideal solution is calculated, using a particular distance function (e.g. Euclidean). One alternative is then preferred to another when it is both closer to the ideal solution, and further away from the worst solution. The best alternative is the one that minimizes the distance from the ideal and maximizes the distance from the worst solution.

Rather than the typical OR optimization methods, which call for continuous sets of alternatives, TOPSIS provides a simple and flexible way to compute preference scores for discrete alternatives. With respect to our decision problem of scoring SRI funds, this seems to be the only appropriate method. TOPSIS provides a really straightforward and easy to understand way of obtaining SRI scores, which can be used in further financial performance research without need for adaptation. The amount of information needed from the decision maker is generally also quite low and no special software packages are required for implementation. However, TOPSIS sometimes yields illogical results, particularly given extreme performances on different criteria. The most appropriate distance function to calculate the overall scores can also be a topic of discussion. For an example of an investment decision application of TOPSIS we refer to Tsao (2003).

4.5 Application of MCDA framework to SRI issues

All of the challenges presented in Section 4.3 can be addressed using MCDA methodologies. As pointed out before, the first three challenges are similar in the sense that they can be met by a social responsibility indicator to score and classify mutual funds. The fourth challenge, which considers the untapped potential of the retail side of the SRI market, requires a MCDA framework that can be tailored to the needs of individual clients to better understand the SRI concept and their social investment preferences.

4.5.1 Social performance indicator

Using the MCDA framework, a social performance indicator for mutual funds can be built. Instead of using a dichotomous classification, an indicator would make it possible to discriminate between mutual funds on the basis of their social performance in a more continuous way. This can be achieved either by calculating a social performance score, or by allocating mutual funds to different categories based on their social performance. Both options can be achieved using MCDA methods. Having a social performance indicator would allow scholars to have an even better look at the corporate social versus corporate financial performance relationship. Regulators could use the indicator to assess the design of SRI funds. Banks can be assisted by such an indicator to improve their line of SRI products. Overall, the indicator would increase the transparency and clarity of the supply of SRI products.

In developing the indicator, it is necessary to first define a relevant and consistent set of criteria for assessing the social responsibility of a mutual fund, taking into account views from different stakeholders. This set of criteria needs to be comprehensive and yet as compact as possible. The assumption of preferential independence also needs to be fulfilled. A useful tool to define the set of criteria is the value-focused thinking approach by Keeney (1992), which structures the process of defining relevant criteria. The definition of a valid set of criteria requires the collaboration of an expert panel representing the different stakeholders involved with the SRI decision process. To help find the relevant groups of stakeholders, Checkland's CATWOE model might be used (Checkland & Scholes, 1990). Applying the AHP to score mutual funds on social performance, Pérez-Gladish and M'Zali (2010) already developed a set of criteria. However, they only used one expert to establish the list of criteria, so future research needs to consult representatives from all different interest groups to ensure that the set of criteria is comprehensive and consistent with the different views from the field of SRI. A first example of a comprehensive set of criteria is displayed in Table 5.2.

Once the consistent set of criteria is found, the alternatives need to be considered. This set of alternatives will depend on the application of the social performance indicator. For academic purposes, a scholar might have a good reason to consider only a small subset of mutual funds. For the purpose of government regulation, the set of alternatives might include all the mutual funds offered in a certain country or region. In applying the social performance indicator in the banking industry, the set of alternatives might consist of different potential mutual funds that the bank wants to release in the near future. Given that the set of alternatives can be changed for the different purposes, it is of keen importance to build the social performance indicator in such a way that it is independent from the underlying set of alternatives. The choice of decision alternatives will also depend on the availability of data and transparency with respect to the investment process. For example, countries like Belgium, the Netherlands and the Scandinavian region have already implemented a great number of transparency guidelines, which makes it easier to measure social responsibility of investment vehicles in those countries.

The next step is to choose a particular MCDA model to build the social performance indicator. From the school of value measurement methods, different methodologies can be applied to the problem at hand. AHP has been successfully implemented in the past, although given a limited number of experts determining the set of criteria. MAVT can also be applied, as it is a more general form of the AHP, on the condition that enough information can be collected from the expert panel about the form of the partial utility functions. Another value measurement method that is feasible for this problem is MACBETH, which is similar to AHP but uses categorical instead of ratio scales. In the outranking school of MCDA methods, both PROMETHEE, together with its visual aid GAIA, and ELECTRE can be used as a tool to classify mutual funds based on outranking relations. The third class of models — goal, aspiration and reference-level methods — is generally less appropriate as a social performance indicator, as it typically requires an association of every criterion with quantitative and measurable attributes, which is not the case with some of the softer criteria related to SRI. However, the TOPSIS reference-level methodology allows for a very simple and efficient calculation of social performance scores without requiring too much quantitative input. The goal programming methods can be used either ex-ante to narrow down a broad set of alternatives, or ex-post to find the optimal way of implementing a fund-of-funds strategy. For all of these methodologies, it is important to consider underlying assumptions (e.g. rank reversal, preferential independence) and to end up with a straightforward quantitative output that can be used in performance regressions. This last condition might be harder to fulfill for the outranking methods.

To implement these different methodologies, and to find the most robust and qualitative one to score mutual funds on their social performance, a lot of information is needed as an input. This information should be collected from the expert panel of representative stakeholders. Therefore, a key success factor is to collect this information in the most efficient way possible, as it seems unlikely that such an expert panel can be consulted regularly. We suggest that a questionnaire, consisting of all the relevant questions to implement the different methodologies, is constructed. This might not be straightforward, as the different methods require quite different sets of information.

Once the criteria have been defined, a sample set of alternatives is constructed and all the necessary information is collected, different methods can be implemented to build the social performance indicator. Instead of choosing only one of the methods ex-ante, we suggest that future research has a look at every feasible model that can be applied to the problem. After the models are built, robustness and sensitivity analyses on these different models can help to determine which model is most qualitative and robust in judging the social performance of a mutual fund. These robustness checks can either be executed using built-in software modules that alter input values and consider the changes in output, or using more statistical procedures like Gini's concept of transvariation (e.g. Van den Bossche, Rogge, Devooght, & Van Puyenbroeck, 2010). More concretely, future research could also look into the stability of obtained SRI scores given changes in the underlying criteria weights. For every criterion, a stability interval can be calculated, indicating to what extent the weight of a criterion can be changed without resulting in a change of SRI rankings. Ideally, the applied method should be highly robust for underlying changes in the weights, as these are very hard to elicit or determine for decision makers, i.e. investors. Furthermore, feedback from the expert panel can be used to determine what kind of model is most in line with the expertise from different stakeholders. Depending on the particular application of the model, several other criteria can be considered in assessing its appropriateness as well. For example, one should always take into account ease of implementation, as the ideal SRI indicator will be implemented on a larger scale both in the professional industry (e.g. asset management) and academia (e.g. performance research). Models that might be more advanced from a technical point of view might be harder to implement because of a greater need for detailed input from the decision maker (e.g. elicitation of utility functions). Also, more advanced methods might be harder to understand by the decision makers, which might complicate the implementation. The MCDA method that provides the best trade-off between robustness from a technical point of view (i.e. has the largest average stability interval), and convenience from a practical point of view (i.e. ease of implementation and understanding) can then be applied as a social performance indicator to address the academic, regulation and commercial bank challenges presented in Section 4.3.

4.5.2 "Green" MiFID

To address the underdeveloped state of the retail end of the SRI mutual fund market, the MCDA framework could also be adopted by banks to help their customers better understand the variety of SRI mutual funds being offered. The main difference with the social performance indicator is that this tool needs to be tailored to the individual needs of investors. Instead of looking for a consensus on the set of criteria and their importance from an expert panel of representative stakeholders, individuals now need to provide input that represents their preferences. Ideally, the best performing model found when constructing the social performance indicator could be used on the individual level as well, just by replacing input from the expert panel by individual information provided by investors. However, chances are that this approach would be too technical for a straightforward implementation in standard investment advisory practice. Therefore, we suggest that future research explores to what extent it is possible to transform the principles behind the best possible MCDA methodology into a standardized questionnaire that would be better suited to help clients in getting to know SRI products and their preferences towards these products in a better way. As an interesting example, we point to the European Commissions' directive called MiFID, which obliges banks to let investors complete a questionnaire to get to know their personal investment profile. As a result, a client interested in investing is sorted into a certain category, which comes with certain rules of protection (Davies et al., 2006). The directive provides transparency and tries to protect retail investors from investing in products that are incompatible with their preferences. In a similar way, we propose a so-called "green" MiFID to be created. Retail investors interested in SRI would need to complete a survey, the questions of which are based on the underlying MCDA framework, to get more insight into their preferences with regard to socially responsible investing.

It is worth noting that this "green" MiFID is compatible with the social performance indicator in opening up the retail side of the SRI market. The social performance indicator at the supply side of the SRI market increases product transparency as it accommodates the comparison of different mutual funds with respect to their performance on ESG criteria. The "green" MiFID provides retail investors with a better insight into their personal SRI profile. Together, these two measures make it possible for individual investors interested in SRI to get a clear understanding of what products are best suited to match their preferences, which exactly addresses the main cause of the SRI retail market remaining underdeveloped.

4.6 Conclusion

We highlight four challenges that are facing SRI and a framework that could be used in future research to address these. A first group of challenges revolves around the need for a social performance indicator that is able to score and/or classify mutual funds with respect to their social responsibility. Such an indicator could enable scholars to better examine the relative performance of SRI mutual funds; it can help the regulatory authorities in developing a certificate for genuinely socially responsible mutual funds; and it accommodates the process for banks to develop new and transparent SRI products. Another challenge concerns the underdevelopment of the retail side of the SRI market. To tap the potential at this side of the market, there is the need for a tool that can help investors to better understand the SRI products and their own preferences with respect to these products.

Given these two groups of challenges, we point to multi-criteria decision analysis (MCDA) as the methodological framework that could be used in future research to develop appropriate solutions. Different MCDA methodologies can be implemented to build a social performance indicator that would address the first group of challenges. Using robustness analyses and feedback from an expert panel representing the main stakeholders in the SRI field, it should be possible to determine the best possible MCDA method to score and classify mutual funds. This would benefit academic research, the regulatory burden of governments and the efforts of banks to create new SRI products. From this MCDA indicator, it should also be possible to create a simple tool based on a standard questionnaire that can help investors to better understand their own preferences with regard to the social responsibility of financial products. Such a questionnaire is comparable to the current MiFID framework that is used in banks to help clients understand their risk preferences when investing in financial products.

Note that the indicator on the one hand, and the "green" MiFID tool on the other hand, are not unrelated. Ideally, they should be aligned as to stimulate the further growth of the SRI market. The indicator, among other things, can create more transparency in the supply of SRI financial products and makes it easier to discriminate between SRI mutual funds. The investor tool can help potential investors to find out about their social performance preferences. Together, these two tools have the potential to open up new perspectives for the retail side of the SRI market, which could further add to the growth of the SRI market.

Chapter 5

Process-oriented social responsibility indicator for mutual funds: A multi-criteria decision analysis approach

Abstract

In recent years, socially responsible investing (SRI) has grown from a niche to a mainstream investment strategy. Consequently, the supply of SRI mutual funds expanded and a discussion on the performance of SRI funds developed. The traditional approach to measuring mutual fund performance does not take into account possible heterogeneity with respect to social responsibility and focuses on a dichotomous SRI versus non-SRI typology. We consider multi-criteria decision analysis (MCDA) to build an indicator to capture the multi-faceted nature of SRI. Considering extensive robustness analysis and additional criteria, we find the PROMETHEE methodology to be most recommended for both academic and professional applications. The obtained scores and rankings prove to be highly robust and are in line with general intuition. The proposed PROMETHEE-based indicator is also easy to understand and implement, which makes it highly suited for implementation in mutual fund performance research.

5.1 Introduction

The practice of socially responsible investing (SRI) has evolved over the past couple of decades. Starting from the moral investing principles of religious organizations, the first generation of SRI was only concerned with excluding companies that did not meet certain ethical standards from the investment universe. This practice is referred to as negative screening. In a later generation, SRI shifted its focus to the application of positive screens, i.e. companies that were performing well according to environmental, social and governance (ESG) issues were selected from the investment universe. The third generation of SRI simply combined the use of positive and negative screens. Shareholder activism is the latest addition to this set of SRI practices. Eurosif (2010) segments the market by means of core versus broad SRI. Core SRI consists of exclusions based on norms and ethical values, and positive screening (incl. best-in-class approach). Broad SRI is composed of simple screening strategies, engagement and integration. The latter category represents the mainstreaming of SRI.

Following several societal challenges like the Vietnam War, the civil rights movement and the racist system of apartheid, the interest in SRI has grown over the years. In the United States, about one out of every nine dollars under professional asset management is now invested in SRI assets (Forum for Sustainable and Responsible Investment, 2012). The volume growth of SRI is also larger than for the conventional asset class. In Europe, we see similar results, with the combined growth of SRI strategies outperforming the conventional market (Eurosif, 2012).

Despite the growth in the SRI market, and the associated maturation from a niche to a mainstream strategy (Sparkes & Cowton, 2004; KPMG & ALFI, 2013), there is no conclusive evidence of SRI funds not underperforming conventional funds. In the field of strategic management, the debate goes back to the opposing views of Friedman (1970) and Freeman (1984), i.e. a shareholder versus a stakeholder view of the firm. Mackey,

Mackey, and Barney (2007) found that the effectiveness of social responsibility crucially depends on the demand and supply conditions for SRI: a finding that reconciles between both extreme views. In financial economics, the literature goes back to the opposing views of Moskowitz (1972) and Markowitz (1952). Moskowitz (1972) published the first empirical paper that indicates there can be financial value to screening strategies. Markowitz (1952), the creator of modern portfolio theory, argues that SRI funds are a subset of the market portfolio and suffer from idiosyncratic risk because of the fewer diversification opportunities that come with the additional constraints to the investment universe. Today, scholars mostly apply asset-pricing regression models to compare risk-adjusted returns between SRI and conventional mutual funds. Notwithstanding the increasing complexity of these methods, results found are often inconclusive and/or weak. The typical approach is to use a multi-factor asset-pricing regression accounting for different sources of risk. Next, the model is applied to a difference portfolio of both SRI and conventional funds, which yields a risk-adjusted return (alpha; constant in the regression). If the constant is significantly different from zero, a conclusion can be drawn on the out- or underperformance of SRI funds.

A main drawback of earlier research is that only typical sources of risk are taken into account, without considering the possible impact of style differences between and among SRI and conventional funds, with respect to the process quality of designing funds based on ESG criteria. Renneboog, Ter Horst and Zhang (2008a) attempt to overcome this issue using an "ethics factor", but do not take into account the multi-dimensional nature of social responsibility. Therefore, the risk-adjusted difference returns found in previous research might be biased. It is reasonable to assume that social responsibility can affect the long-term risk faced by corporations. For example, when a firm is investing in renewable energy, it is less likely to be negatively impacted by an energy crisis in the future. A better relationship with the employees can help reduce the chance of future strikes. Social responsibility should thus be taken into account as a factor when examining the performance of mutual funds. Considering the specific social responsibility content of individual firms is unfeasible at the aggregate level of a mutual fund. Alternatively, we build a proxy indicator for social responsibility by measuring the quality of the design process of putting together SRI mutual funds. Such a SRI process quality indicator can be used to help discriminate between mutual funds with respect to process-oriented social responsibility, instead of the traditional dichotomous approach (SRI vs. non-SRI), which does not allow researchers to account for heterogeneity among both the group of SRI and conventional funds. This can add more detailed insights into the risk-adjusted performance relation between the different types of funds. Next to the academic application, these scores can be insightful to address professional challenges as well. For example, managers could use such a tool when constructing new investment funds.

Different studies have tried to measure a mutual fund's social responsibility in the past. Most commonly, several papers have described the development of tools for individual investors to take into account both financial and ESG perspectives. Hallerbach, Ning, Soppe and Spronk (2004) present an interactive multiple goal programming approach to help investors compose SRI portfolios from individual assets. Cabello, Ruiz, Pérez-Gladish and Méndez-Rodriguez (2014) complement this research by proposing a reference point method for scoring SRI mutual funds with respect to environmental performance. Despite its usefulness for individual investors in making better investment decisions, these kinds of approaches are rendered ineffective for implementation in further SRI performance research given the required number of subjective assessments by the decision maker. These approaches are also less instrumental for fund managers looking to optimize the design of new funds. We focus on SRI performance research and possibilities to take into account heterogeneity among funds with respect to ESG criteria. Additionally, we consider to what extent this kind of method can be used in a more professional setting to improve the design of mutual funds with respect to social responsibility and to help investors choose an SRI fund. Ideally, this calls for techniques that are easy to implement and replicate and thus do not require any input from experts or individuals to assess criteria or elicit preferences. One way to achieve this goal is to simply consider the number and variety of applied screens to approximate social responsibility (e.g. Renneboog, Ter Horst, & Zhang, 2008b). However, this approach reduces the concept of SRI to merely screening activities. To build a process quality indicator that reflects the multiple dimensions to the concept of SRI, we therefore consider the field of multi-criteria decision analysis (MCDA). Pérez-Gladish and M'Zali (2010) and De Moor, Devooght and De Bondt (2012) were the first to design an indicator to measure the process-oriented social responsibility of mutual funds. However, they only implemented the AHP methodology and disregarded other possibly superior methodological approaches. In contrast, we consider the entire set of established MCDA techniques. From an extensive analysis of the different methods within the MCDA framework, four eligible methods are applied. To find which of the four considered MCDA methods performs best to approximate social responsibility in a way that can be useful for future SRI performance research, we conduct extensive robustness analysis and take into account additional criteria depending on the type of application. We also implement the indicator on a larger sample in a case study of the Belgian SRI fund market.

5.2 MCDA methodologies

5.2.1 Background theory of MCDA

Multi-criteria decision analysis (MCDA) is an operations research tool to help solve delicate decision problems that involve multiple conflicting decision factors (Figueira, Greco, & Ehrgott, 2005). Within the MCDA paradigm, there often is no such thing as the optimal solution. Rather, a decision maker evaluates a set of alternatives considering all decision factors or criteria that are relevant to the analysis. The main difficulty in the decision process is the aggregation of relevant criteria. The specificity of the different methods within MCDA lies mostly in the way this aggregation problem is addressed. Given the multiple dimensions involved when assessing the social responsibility of mutual funds, and the need for subjective judgments, MCDA provides an interesting framework to build a SRI design process quality indicator. The transparency of the MCDA framework can help decision makers (e.g. fund managers) to better understand how the social performance scores were calculated. All in all, the MCDA tools can help to objectivize some of the subjective information when assessing social responsibility.

Every MCDA method consists of five building blocks (Belton & Stewart, 2002). In a first step, a decision maker determines the consistent family of relevant criteria (Roy, 1974). Next, alternatives considered to the decision problem are introduced. Then, every criterion is scored with respect to each alternative to construct a performance table. To determine the relative importance of the different criteria, weights need to be elicited. This can be done in a number of ways (e.g. directly or via pairwise comparison). Finally, the performance table is transformed into a score using approaches that are specific to the different MCDA methods.

The simplest approach to address multi-criteria problems is a weighted sum (Figueira et al., 2005). However, this requires the often-untenable assumption of linearity of preferences. To better accommodate complex decision processes, a plethora of methods has been developed in the MCDA field. These methods can be divided into three schools of thought: value measurement methods, outranking methods and goal, aspiration or reference level methods. Assuming that a bad score on one criterion can be compensated by a good score on another criterion, value measurement methods evaluate a global score for every alternative based on partial scores with respect to the different criteria. Outranking methods allow for the notion of incomparability, i.e. partial scores cannot be compensated, and are based on pairwise comparisons. Using feasibility constraints, goal, aspiration or reference level methods can allocate goals to every criterion to identify the most desirable alternative.

For the purpose of our process quality indicator, we examine the established MCDA methods. To find which methods are feasible to our decision problem, we consider different criteria that apply to our ideal process quality indicator. First, we require a quantitative output for the indicator, as we want to use the score in asset pricing regressions. The method also needs to be state-of-the-art. Additionally, we need to be able to deal with qualitative inputs, not just hard quantitative inputs. The criteria used in assessing the process quality should be easy to translate into the model. Furthermore, the method is ideally easy to implement, using either existing software or by programming in statistical software. Finally, we want to limit the amount of expert judgments required to calculate the scores. The comparison of the different MCDA methods with respect to these criteria can be found in Table 5.1.

Comparisons	**Score output**	**State-of-the-art**	**Qualitative input**	**Translation of criteria**	**Implementation**	**Judgments**
			Value measurement methods			
Analytic hierarchy process	x	x	x	x	x	x
Analytic network process	x	x	x	x	x	
MACBETH	x	x	x	x	x	x
Multi-attribute utility theory	x	x	x	x	x	x
UTADIS/GRIP	x	x	x	x	x	
			Outranking methods			
ELECTRE			x	x		
PROMETHEE	x	x	x	x	x	x
			Goal, aspiration or reference level methods			
TOPSIS	x	x	x	x	x	x
Goal programming	x	x			x	x
Data envelopment analysis*	x	x			x	x
			Other			
Decision rules		x	x	x	x	

The different methods from the field of MCDA assessed with respect to the ideal characteristics of the process-oriented social responsibility indicator.

Table 5.1: *MCDA method comparison*

Note that we also included data envelopment analysis (DEA) as a possible method, even though it does not really belong in the MCDA universe. From the analysis, we find five methods to be highly feasible: analytic hierarchy process (AHP), MACBETH, multi-attribute utility theory (MAUT), PROMETHEE and TOPSIS. Multi-attribute utility theory and PROMETHEE normally call for elaborate value judgments, but the specific definition of our decision criteria makes them feasible to our problem as well (see Section 5.2.2). The value measurement methods seem most appropriate to build a process quality indicator. Specifically, we consider the analytic hierarchy process (AHP), MACBETH and multi-attribute utility theory (MAUT). From the outranking school, we implement the PROMETHEE II ranking methodology, which yields a complete ranking of alternatives based on net flow scores, which can be used as a quantitative input in asset pricing regressions to control for differences in social responsibility. Goal, aspiration or reference level models are generally less well suited as most of the methods call for highly quantifiable criteria and are mostly useful when the set of alternatives is continuous and infinite rather than discrete and limited. One exception is the simple TOPSIS methodology, which we apply as well. From a thorough robustness analysis and additional criteria depending on the type of application of the scores, we are able to determine which specific MCDA method is most recommended to serve as a design process quality indicator in future research.

Before introducing the different MCDA methods we implement to find a social responsibility proxy score, we first present two building blocks needed in every MCDA exercise: the relevant criteria, and the decision alternatives considered.

5.2.2 Criteria

In building an MCDA indicator, we need to define assessment criteria. Following Roy (1974), a consistent family of criteria satisfies the axioms of exhaustibility, cohesion and non-redundancy. Pérez-Gladish and M'Zali (2010) conduct extensive literature research to come up with a hierarchy that is split up between criteria associated to the actual content of a SRI fund on the one hand, and the transparency and credibility of the SRI fund on the other hand. A major drawback of their study is the need for expert judgment in comparing different alternatives, which makes it harder to replicate the methodology in other research. De Moor et al. (2012) use the same split of criteria, but formulate them in such a way that information can simply be extracted from publicly available information about the fund (e.g. prospectus), without need for extra judgment from the decision maker, who now only needs to provide preference information about the importance of criteria. As we aim to build a process quality indicator that can easily be implemented in future financial economics research, we start from the criteria hierarchy of De Moor et al. (2012). Then we omit three criteria[1] because of redundancy, generalize another criterion[2] and add one complementary criterion[3] that was pointed out by Pérez-Gladish and M'Zali (2010), the Belgian financial sector organization Febelfin (2012) and the United Nations (2013) responsible investment initiative.

The eventual hierarchy (Table 5.2) consists of different statements about the process-oriented social responsibility of a mutual fund. Note that all of these criteria concern the social responsibility content of the design process of how a mutual fund was constructed, rather than the company-specific social responsibility content. Hence, we refer to our indicator as process-oriented. Either a fund complies with the criterion or it does not. As a consequence, the performance table is dummy-coded and only requires publicly disclosed information without need for any expert judgment. Even though this constraint on the performance table limits the power of several MCDA methods to some extent, this approach towards the criteria greatly

[1] Assessment by means of controversies, exclusion of disputable technologies and active and regular audit of investments.

[2] Investment is principally ($> 75\%$) in companies that invest in sustainable technologies (not just sustainable electricity and CO_2 reduction).

[3] Incorporation of SRI principles established by reputable organizations.

contributes to the replicability and applicability of the indicator in future research. An opposite approach, which allows for more elaborate judgment, can lead to richer results, but would not be feasible for implementation in performance evaluation research, as it would require an infeasible need for expert judgment. As our eventual goal is to shed new light on the performance relation between SRI and conventional funds, we opt for the more restrictive approach. The dummy coding also simplifies the process of eliciting value and preference functions, which solves the need for more elaborate judgment in multi-attribute utility theory and PROMETHEE. This in turn means that we have five feasible methods to build a process indicator from. Because of the MCDA aggregation and comparison across 20 different criteria, using the criteria weights, the scoring output is not dichotomous, which still enables us to capture the heterogeneity between different mutual funds with respect to the quality of the SRI screening process.

5.2.3 Alternatives

Next to the criteria, every MCDA method also calls for specific alternatives to be defined. In the case of a SRI process indicator, the ideal set of alternatives consists of every mutual fund available to an investor. However, it would be unfeasible to gather the necessary data to apply the MCDA methods to such a large set of alternatives. Instead, we opt for a small sample set of alternatives, as the focus of this paper is on the indicator methodology. We simply illustrate the application of the indicator to a limited set of alternatives from SRI mutual fund providers on the Belgian market. The choice for the Belgian market is motivated by the availability of an official list of recognized sustainable products by the Belgian financial industry organization (Febelfin), which is accompanied by additional documentation on the mutual fund design process for the different SRI providers active on the Belgian market. This kind of information simplifies the completion of the performance table to a great extent. Specifically,

Goal	Criteria	Subcriteria	Sub-subcriteria
Social performance indicator	Screening process and consistency	C1. Priority screening process	
		Independent data gathering and analysis of sustainability	C2. Data gathering and analysis of sustainability by independent external specialists (e.g. EIRIS) C3. Incorporation of SRI principles established by reputable organizations (e.g. UN SRI, Febelfin) C4. Information from stakeholders and relevant NGOs
		Positive selection criteria	C5. Best-in-class approach with respect to ESG criteria C6. Use of sector specific positive criteria C7. Investment is principally (>75%) in companies that invest in sustainable technologies
		Negative selection criteria	C8. Use of categorical rejects C9. Assessment by means of negative criteria C10. Exclusion of unsustainable technologies
		Monitoring and updates	C11. A research teams checks legal and regulatory developments, trends and behavior of companies such that criteria are in line with recent societal developments C12. Monitoring if portfolio is consistent with defined criteria (continuously, sector specific or occasion specific)
		Dialogue	C13. Companies are informed about conclusions selection methodology C14. Active engagement policy (constructive and critical dialogue with companies in portfolio) C15. Active voting policy (voting at companies' shareholder meetings)
	Transparency and control	Transparency	C16. Release of qualitative information about the screening process (e.g. applied screens) C17. Release of quantitative information about the screening process (e.g. scores) C18. Release of current portfolio C19. Compliance with external transparency guidelines (e.g. Eurosif/Belsif)
		C20. Board of experts	

Schematic presentation of the hierarchy of criteria, with the 20 bottom-level criteria used in the analysis. An in-depth presentation of the different criteria can be found in Appendix 5.A.

Table 5.2: *Hierarchy of criteria*

we consider 9 alternatives, which can be divided into two categories according to whether the fund is labeled as an SRI fund or as a conventional fund by its issuer (Table 5.3). Larger and more international samples can be considered in the future, upon availability of information on the SRI investment process.

SRI funds	**Conventional funds**
BNPP L1 Equity World Aqua (LU0831546592)	BNPP L1 Equity World (LU0072778490)
Dexia L Sustainable World (LU0113400328)	Dexia Quant Equities World (LU0235267860)
ING L Invest Sustainable Equity (LU0394658412)	ING L Invest World (LU0119219730)
KBC EcoFund World (BE0133741752)	KBC Equity Fund World (BE6213775529)
Triodos Sustainable Pioneer Fund (LU0278272843)	

Presentation of the funds considered in the analysis, including the respective ISIN-code.

Table 5.3: *Alternatives*

For the four largest commercial banks active in Belgium, we include both a SRI fund and a matching (e.g. same international orientation) conventional fund. Additionally, we include a fund from Triodos, which is a commercial bank that specifically focuses on sustainable finance. All of the funds are capitalization funds, are geared towards equity and are oriented towards the global market.

The current sample should be considered a first example from a well-developed SRI market that is easy to investigate because of highly transparent reporting. In Section 5.4, we include results from a larger sample as well, but considering only one methodology, i.e. the one we find to be most appropriate for this research question. Future research should look into the further application of the scoring tool on more and other alternatives, which will be possible when more funds start to transparently communicate on their underlying SRI investment process, rather than just releasing mostly uninformative legal documents. The only element of the analysis that changes is the performance table. To limit the data gathering effort, data envelopment analysis (DEA) could also be applied a priori to a broader set of alternatives to bring it down to a more limited set that is more feasible for the MCDA indicator.

5.2.4 Feasible methodologies

After establishing criteria and alternatives, we now need to find an appropriate aggregation method to calculate an overall SRI process quality score for each mutual fund. We specifically aim for a scoring output, as we want to include this score in future SRI performance research. For other purposes, like the development of an SRI tool for regulatory organizations, it might be wiser to search for a classification, rather than a scoring/ranking output.

The different MCDA methods use alternative approaches to address the aggregation problem. We consider five methods, which are suited for the SRI performance indicator problem: analytic hierarchy process (AHP), MACBETH, multi-attribute utility theory (MAUT), PROMETHEE and TOPSIS. Next, we describe each of these five methods, why they are useful for the purpose of a process quality indicator and how they are implemented. More technical details are presented in Appendix 5.D.

Analytic hierarchy process

First presented by Saaty (1977, 1980), the analytic hierarchy process (AHP) is an elegant approach for dealing with discrete alternative problems with multiple conflicting criteria. The technique can be classified as a value measurement method. In essence, AHP consists of three steps. In a first step, the problem needs to be structured by building a hierarchy of criteria relevant to the nature of the problem at hand. On top of the hierarchy, the goal of the AHP analysis is stated. Below the goal, criteria are listed. Depending on the problem, it might be necessary to create additional levels of subcriteria and sub-subcriteria as well. The final level in the hierarchy consists of the alternatives considered. An important assumption for each of the methods we implement is the independence of criteria. The AHP hierarchy for the SRI process indicator looks like the criteria in Table 5.2, with the addition of the alternatives at the end of each of the 20 bottom-level criteria. The same hierarchy is also used for the other methods implemented in this paper.

The second step of the AHP is the scoring process, which leads to a priority ranking of alternatives and criteria. Alternatives need to be scored with respect to all the lowest-level criteria in the hierarchy. In our case, we define the criteria in such a way that alternatives receive a true (1) or false (0) quote for every criterion, as was explained before. For example, if we find in a fund's publicly disclosed information that the screening process is given priority over the financial analysis, the fund receives a true (1) score for that criterion. The eventual performance table can be found in Appendix 5.B. Next to scoring the alternatives with respect to the criteria, priorities need to be calculated. In the AHP this is typically done by pairwise comparison of the bottom-level criteria within the same group using a verbal importance scale from 1 to 9. Psychologists argue that this technique is easier for decision makers when expressing their preferences with regard to the importance of different criteria. Instead of allocating global weights to the criteria in the hierarchy, AHP guides the decision maker through several pairwise comparisons of bottom-level criteria from the same subgroup, indicating which criterion is preferred and to what extent. All of these pairwise comparisons are then combined into a comparison table, which is then checked for consistency. When the table is deemed consistent the criteria priorities and overall priority scores can be used in SRI performance research. An overview of the questionnaire including the AHP comparisons presented to the two experts to elicit weights can be found in Appendix 5.C.

In a last step of the AHP, a robustness analysis is usually conducted to examine how much the results change when inputs (e.g. weights) are changed. This is an important step to determine the reliability of results. Note that the results from one of the experts are used to conduct our initial analysis. Results from the second expert are used to test for overall robustness, in addition to the analysis of stability intervals. To implement the AHP technique in building a SRI process indicator, we make use of the Expert Choice software package.

MACBETH

MACBETH, which stands for "Measuring Attractiveness by a Categorical Based Evaluation Technique", is very similar to the AHP and can also be considered a value measurement approach (Bana e Costa, De Corte, & Vansnick, 2005). The first step again requires a hierarchy of criteria. We use the same hierarchy as for the AHP (Table 5.2).

In the second step, the scoring process needs to be completed. Like for the AHP, the alternatives can be evaluated directly for each criterion, using publicly available information about a fund. MACBETH also requires the decision maker to indicate the appreciation for an increase from false (0) to true (1) for every criterion using a seven-point semantic scale. Next, criteria are weighted using pairwise comparisons. The difference with AHP, however, is in the scale used to make the comparisons. Instead of a ratio scale, MACBETH uses an interval scale. In making the comparisons, MACBETH requires the decision maker to indicate preferences on a semantic scale with seven categories. Unlike AHP, MACBETH also requires the pairwise comparison of every bottom-level criterion, and not just those bottom-level criteria within the same subcategory. For the process quality indicator, this greatly increases the number of judgments needed, which adds complexity to the assessment. Instead, we simply consider the weights found by means of the AHP approach. This should not depreciate the value of the overall findings, given the extensive robustness analysis further on in this paper. The proposed AHP weights are then converted on a MACBETH scale using linear programming. This approach also circumvents potential compatibility problems, as MACBETH requires a very high level of consistency in the judgment table to be able to calculate the attractiveness levels for all the alternatives. From the performance table of the alternatives and the MACBETH weights calculated from the AHP approach, the overall attractiveness of the alternatives can be calculated. This measure of attractiveness can be interpreted as the SRI design process quality score.

In a final step sensitivity and robustness checks can again be performed. The focus is primarily on the weight of criteria and the impact on the final scores. Both the weights from a second expert and stability intervals are used to perform this robustness check. We implement the MACBETH methodology using the M-MACBETH software package.

Multi-attribute utility theory

Multi-attribute utility theory (MAUT) is a value measurement method that distinguishes itself by the use of utility functions to transform the performance table (Keeney & Raiffa, 1976). The analysis starts again from a hierarchy of criteria and the scoring of considered alternatives with respect to these criteria. Next, the marginal scores for every criterion are transformed into utilities using marginal utility functions that need to be elicited. This is typically the most important and difficult step in MAUT. To find the marginal utility function for every criterion, a lot of information is needed from the decision maker. Given our definition of criteria as true or false statements, however, this elicitation process is simplified to a great extent. Since every criterion can only receive one of two scores (1 or 0), the utility function is a simple two-point discrete function that needs no further information for elicitation.

Next to the elicitation of utility functions, the decision maker needs to specify the weights of the criteria. There is no built-in way of doing this in MAUT, and thus alternatives need to be considered. We again use the weights found from the AHP questionnaire and further examine robustness in Section 5.4. Once the marginal utility functions and weights are established, aggregation is needed to find the global utility scores by which the alternatives can be ranked. Typically, an additive aggregation model is used, although alternatives are also possible (e.g. geometric). The obtained scores serve as the proxy for the SRI process quality of the mutual funds. Note that the weighted sum approach is nested within MAUT when all marginal utility functions are assumed to be linear.

Because of the specific definition of our criteria as individual statements that are evaluated dichotomously, the added value of MAUT in using underlying utility functions disappears. The dummy-coded criteria require a simple discrete utility function, which yields a utility table that looks exactly like the original performance table. Since the weighting process is also not specific to MAUT, but taken from the AHP approach, results found by means of MAUT are exactly the same as for AHP. Consequently, we do not include the results from MAUT in the next sections, as in our specific case, MAUT is a special case of AHP.

PROMETHEE

PROMETHEE stands for "Preference Ranking Organization METHod for Enriched Evaluation" and belongs to the outranking approach. It was originally developed by Brans and Vincke (1985) and can be somewhat considered as an improvement over the ELECTRE outranking techniques that originated with Roy (1985). A hierarchy of criteria, the alternatives considered to the problem and their performances are again required at the start of the process. Then, pairwise preference degrees between alternatives are computed by means of a preference function. A preference degree expresses how the decision maker prefers one alternative to another for a certain criterion, based on an underlying preference function that needs to be decided on. This kind of unicriterion performance comparison between alternatives is typical for the outranking approach. The elicitation of the preference function is again straightforward, as our definition of the criteria leads to a discrete preference function that facilitates the PROMETHEE analysis. Given the preference degrees, unicriterion positive, negative and net flows are computed, which summarize how every alternative is preferred to other alternatives. Finally, global net flows are computed by combining the unicriterion net flows, using the criteria weights obtained from the SRI experts. From the global net flows, PROMETHEE II is used to construct a complete ranking of the alternatives. The obtained net flows can be used as a quantitative input to control for differences in social responsibility, just like for the value measurement methods considered in our research.

As for MAUT, there is no built-in way of defining the weights of the different criteria. Therefore, we again implement weights obtained from the AHP questionnaire, which can then be used to aggregate the unicriterion flows to global net flows per alternative. These global net flows are the quantitative output from PROMETHEE that can be used as a proxy for the process quality indicator.

An interesting feature of PROMETHEE is that it can be used in conjunction with the GAIA (Geometric Analysis for Interactive Aid) visualization procedure, which presents the results in a two- or three-dimensional plane. GAIA not only presents results, but can also be used to enrich the sensitivity analysis to test for the quality of the obtained results. The PROMETHEE method is implemented by means of the D-Sight software.

TOPSIS

TOPSIS is the acronym for "Technique of Order Preference Similarity to the Ideal Solution" and is part of the third class of MCDA methods, i.e.

goal, aspiration and reference-level approaches (Behzadian, Otaghsara, Yazdani, & Ignatius, 2012). As for every other method, the technique starts with defining alternatives and criteria, and introducing the performance table. The only subjective parameter needed in TOPSIS is the elicitation of the weights, for which there is no built-in method. Again, we implement weights found from the experts through the AHP questionnaire. From a weighted version of the performance table, TOPSIS constructs a ranking based on the ratio of the Euclidean distances of every alternative to the ideal (true statement for every criterion) and anti-ideal solution (false statement for every criterion). This ratio is a normalized score between 0 and 1, which can be interpreted as a process quality score. The best alternative will be the one that maximizes the distance from the anti-ideal solution and minimizes the distance from the ideal solution at the same time. Changing the weights on the different criteria can help determine the robustness of the ranking and results. We implement the TOPSIS analysis by simply programming the method in a spreadsheet.

Most of the methods based on a goal programming, aspiration or reference level approach are not suited for the purpose of our process quality indicator. By including TOPSIS, we include methods from all three schools of MCDA, which greatly contributes to the representativeness of our search for the best performing SRI process indicator.

5.3 Results

For each of the four considered methods the same routine is used to obtain scores on the process-oriented social responsibility of a mutual fund. We first introduce the criteria and the alternatives, which were presented earlier. Next, the performance table (Appendix 5.B), combining criteria to the different alternatives, is filled in by examining publicly disclosed fund information (e.g. prospectus and transparency documents). Then we introduce the weights found from one of the experts filling in the AHP-type questionnaire. In the next section we consider the weights found from the other expert as a way of examining robustness of results. The performance table, together with the weights, is then transformed by means of the four specific MCDA methods. From this transformation, we find our process-oriented sustainability scores, which are presented in Table 5.4.

In the top panel of Table 5.4 we present the scores as we find them from the different methods. A first important observation is that the ranking of the funds is completely consistent across the four methodologies. On top of the ranking we find the theoretical alternative for which the performance table is maximized, i.e. the alternative that receives a "true" statement for every criterion. This theoretical alternative is what other funds should consider in order to improve their current social responsibility score. The fund by Triodos is found to be the most sustainable alternative from a process point of view. This result should not be surprising, as Triodos is a niche player in the banking industry focusing specifically on sustainability and social responsibility. The rest of the ranking is also in line with generally accepted intuition in the Belgian financial industry. Behind Triodos, we first find the SRI funds by the four traditional banks active on the Belgian market, followed by the conventional funds. An interesting observation is that the ranking among the traditional banks is inverted going from SRI to conventional funds. For SRI funds, KBC and Dexia seem to have the most thorough design methodology. However, BNP Paribas Fortis and ING are able to better design their conventional funds with respect to social responsibility, in contrast to KBC and Dexia.

One issue with the scoring output from the four methodologies is the difference in scaling. Therefore, we present the normalized scores in the bottom panel of Table 5.4. Scores are obtained by dividing the original score by the score of the theoretical optimal alternative for the respective methodology. With respect to ranking, the exact same insights are found as for the top panel of Table 5.4. Additionally, we now get some grasp of the relative differences in scores between the alternatives over the different method-

Summary results (non-normalized)				
Alternatives	*AHP*	*PROMETHEE*	*TOPSIS*	*MACBETH*
Max	0.158	0.7016	0.082050188	1
Triodos Sustainable Pioneer Fund	0.15	0.6739	0.08144545	0.9336
KBC EcoFund World	0.139	0.6376	0.07966546	0.8622
Dexia L Sustainable World	0.138	0.632	0.07963925	0.852
ING L Invest Sustainable Equity	0.128	0.6167	0.079327669	0.8061
BNPP L1 Equity World Aqua	0.125	0.5861	0.078209529	0.7143
BNPP L1 Equity World	0.052	0.3295	0.030019509	0.4336
ING L Invest World	0.049	0.3175	0.029618428	0.3979
KBC Equity Fund World	0.032	0.2585	0.022141209	0.2857
Dexia Quant Equities World	0.029	0.2466	0.021608729	0.25

Summary results (normalized)				
Alternatives	*AHP*	*PROMETHEE*	*TOPSIS*	*MACBETH*
Max	100.00%	100.00%	100.00%	100.00%
Triodos Sustainable Pioneer Fund	94.94%	96.05%	99.26%	93.36%
KBC EcoFund World	87.97%	90.88%	97.09%	86.22%
Dexia L Sustainable World	87.34%	90.08%	97.06%	85.20%
ING L Invest Sustainable Equity	81.01%	87.90%	96.68%	80.61%
BNPP L1 Equity World Aqua	79.11%	83.54%	95.32%	71.43%
BNPP L1 Equity World	32.91%	46.96%	36.59%	43.36%
ING L Invest World	31.01%	45.25%	36.10%	39.79%
KBC Equity Fund World	20.25%	36.84%	26.98%	28.57%
Dexia Quant Equities World	18.35%	35.15%	26.34%	25.00%

The top panel presents the results as they are found from the different methodologies. We included one theoretical alternative that indicates the maximum attainable score. The bottom panel presents normalized results with respect to the theoretical maximum.

Table 5.4: *Summary of results*

ologies. We notice that the Triodos fund scores relatively well and is very close to the theoretical maximum. However, the difference between the ideal and the Triodos fund somewhat differs between the methods. From the TOPSIS-ranking, the Triodos fund is less than one percent away from the optimum. For the other methods, this difference is slightly larger. The same kind of relative differences are found for the remainder of the ranking as well. Apart from some variation in relative performance, general insights and ranking of the alternatives is completely consistent over the four methodologies.

As indicated before, the scores found from the different methods can be instrumental in a number of ways. First, a ranking can be made like in Table 5.4, which tells us something about the relative performance with respect to social responsibility of different funds provided by financial institutions. Second, the scores can be used in academic applications as an extra quantitative input in performance research (e.g. asset pricing regressions, DEA). Finally, financial institutions can implement the scores and the applied methodologies in a professional context to improve the design process of their funds with respect to social responsibility.

5.4 Discussion

From the available set of MCDA methods and the three schools of thought, we implemented four eligible methodologies, which led to consistent and identical rankings (Spearman's ρ of 1). Now, we aim to determine which specific method is most appropriate as a process quality indicator. For every application, we require that the obtained scores are as robust as possible. Other criteria depend on the type of application. First, we consider the application of the obtained scores to further examine the performance of SRI funds (academic application). To this end, we also take into account ease of implementation and understanding of the methodology. Second, we also examine what methodology is most recommended to help banks optimize the design of a fund with respect to social responsibility and to help clients select an SRI fund (professional application). Here, we consider the extent to which sensitivity analyses can be performed. A summary of these considerations can be found in the application comparison table (Table 5.5).

Application comparisons	**AHP**	**MACBETH**	**PROMETHEE**	**TOPSIS**
Overall robustness		x	x	x
		Academic application		
Ease of implementation	x		x	x
Transparency and ease of understanding	x		x	x
		Professional application		
Extensiveness of sensitivity analysis			x	

Comparison of the applied methods with respect to the criteria relevant to both the academic and professional application of the indicator.

Table 5.5: *Application comparison table*

Overall robustness of results is obviously an important criterion for both types of application. To comment on robustness, we calculate stability intervals, which indicate between which extreme values the weights of a particular criterion can change such that the best alternative still remains on top of the ranking. Also, we consider the weights found from a second independent SRI expert. Stability intervals and weights for each criterion with respect to the four considered methods are presented in Table 5.6.

For every criterion, we first indicate the allocated weight given the responses of the first and second expert on the AHP questionnaire. Next, we present the lower- and upper bound between which criteria weights can change in order for the best alternative to remain on top of the ranking. Note that for MACBETH the AHP weights are converted on the MACBETH interval scale using linear programming. Therefore, we included the transformed MACBETH weights as well in two separate columns next to the MACBETH stability intervals. For example, the first criterion receives a weight of 33.15% from the first expert and 2.40% from the second

Bottom-level criteria weight robustness												
Criterion	*Weight E1*	*Weight E2*	*AHP*		*PROMETHEE*		*TOPSIS*		*Weight E1*	*Weight E2*	*MACBETH*	
			L	U	L	U	L	U			L	U
Priority screening process	33.15%	2.40%	0.00%	95.50%	0.00%	100.00%	0.00%	100.00%	8.67%	3.48%	0.00%	100.00%
Data gathering and analysis of sustainability by independent external specialists (e.g. EIRIS)	2.16%	23.61%	0.00%	95.50%	0.00%	100.00%	0.00%	100.00%	3.57%	9.45%	0.00%	100.00%
Incorporation of SRI principles established by reputable organizations (e.g. UN SRI, Febelfin)	5.03%	3.27%	0.00%	93.80%	0.00%	100.00%	0.00%	100.00%	6.12%	4.48%	0.00%	100.00%
Information from stakeholders and relevant NGOs	0.56%	6.80%	0.00%	94.90%	0.00%	100.00%	0.00%	100.00%	2.04%	7.96%	0.00%	100.00%
Best-in-class approach for criteria with respect to ESG criteria	5.74%	9.20%	0.00%	94.10%	0.00%	100.00%	0.00%	100.00%	7.65%	8.95%	0.00%	100.00%
Use of sector specific positive criteria	2.76%	2.13%	0.00%	94.40%	0.00%	100.00%	0.00%	100.00%	4.59%	2.99%	0.00%	100.00%
Investment is principally (>75%) in companies that invest in sustainable technologies	0.49%	4.43%	0.00%	3.00%	0.00%	7.08%	0.00%	7.08%	1.02%	5.97%	0.00%	7.62%
Use of categorical rejects	7.57%	2.47%	0.00%	93.90%	0.00%	100.00%	0.00%	100.00%	8.16%	3.98%	0.00%	100.00%
Assessment by means of negative criteria	7.57%	5.14%	0.00%	95.60%	0.00%	100.00%	0.00%	100.00%	8.16%	6.47%	0.00%	100.00%
Exclusion of unsustainable technologies	7.57%	3.56%	6.40%	100.00%	0.00%	100.00%	0.00%	100.00%	8.16%	4.98%	0.00%	100.00%
A research teams checks legal and regulatory developments, trends and behavior of companies such that criteria are in line with recent societal developments	5.25%	1.97%	0.00%	95.60%	0.00%	100.00%	0.00%	100.00%	6.63%	2.49%	0.00%	100.00%
Monitoring if portfolio is consistent with defined criteria (continuously, sector specific or occasion specific)	1.75%	5.91%	0.00%	93.20%	0.00%	100.00%	0.00%	100.00%	3.06%	6.96%	0.00%	100.00%
Companies are informed about conclusions selection methodology	0.24%	1.93%	0.00%	93.60%	0.00%	100.00%	0.00%	100.00%	0.51%	1.99%	0.00%	100.00%
Active engagement policy (constructive and critical dialogue with companies in portfolio)	0.51%	0.93%	0.00%	93.50%	0.00%	100.00%	0.00%	100.00%	1.53%	0.50%	0.00%	100.00%
Active voting policy (voting at companies' shareholder meetings)	2.83%	1.34%	0.00%	93.60%	0.00%	100.00%	0.00%	100.00%	5.11%	1.00%	0.00%	100.00%
Release of qualitative information about the screening process (e.g. applied screens)	4.00%	3.90%	0.00%	93.30%	0.00%	100.00%	0.00%	100.00%	5.62%	5.47%	0.00%	100.00%
Release of quantitative information about the screening process (e.g. scores)	4.00%	3.90%	0.00%	100.00%	0.00%	100.00%	0.00%	100.00%	5.62%	5.47%	0.00%	100.00%
Release of current portfolio	2.23%	9.13%	0.00%	93.60%	0.00%	100.00%	0.00%	100.00%	4.08%	8.45%	0.00%	100.00%
Compliance with external transparency guidelines (e.g. Eurosif/Belsif)	0.91%	1.80%	0.00%	93.20%	0.00%	100.00%	0.00%	100.00%	2.56%	1.50%	0.00%	100.00%
Board of experts	5.56%	6.25%	0.00%	93.50%	0.00%	100.00%	0.00%	100.00%	7.14%	7.46%	0.00%	100.00%
Summary	***4.97%***		**89.87%**		**95.35%**		**95.35%**		***2.81%***		**95.38%**	

Presentation of the stability intervals, between lower-bound (L) and upper-bound (U), for each of the 20 bottom-level criteria with respect to the four applied methods. The average width of the intervals for every method is calculated on the bottom of the table, as well as the average absolute distance between weights from the two experts (bottom of column 1 and 4).

Table 5.6: *Stability intervals*

expert (8.67% vs. 3.48% for the MACBETH weights). Apparently, there is a strong lack of consensus between the experts with regard to this criterion. However, the stability interval for the AHP method goes from 0% to 95.5%; the other methods have a complete stability interval between 0% and 100%. Concretely, this means that the fund from Triodos bank will remain the best performing fund when the weight allocated to the first criterion changes between 0% and 95.5% (resp. 100%). If the weight should become 95.6% or higher, the KBC SRI fund will come on top of the ranking for the AHP methodology. For the other methodologies, the top of the ranking will be insensitive to changes in the weight of the first criterion. To summarize the information from the numerous stability intervals for every methodology across the 20 criteria, we calculate the average distance between the lower- and upper bound. AHP seems to be least robust with an average interval width of 89.87%, which is clearly lower than for the other methodologies. PROMETHEE and TOPSIS have an average interval width of 95.35%. For MACBETH, the average width is 95.38%. Note that we discard this very small differential between PROMETHEE and TOPSIS, and MACBETH. Overall, we could conclude that AHP clearly underperforms the other methodologies with respect to robustness. PROMETHEE, TOPSIS and MACBETH can be considered "most robust" and thus meet the first criterion for both the academic and professional application of the indicator.

If we look at the stability intervals for PROMETHEE, TOPSIS and MACBETH in more detail, we see that the interval width is 100% for every criterion, except one. Apparently, the weight on the criterion determining whether the fund invests principally in companies that invest in sustainable technologies is really driving results. However, considering the judgments from our two individually consulted experts, there seems to be no impact on the ranking whatsoever. It might be of interest to further investigate the role and importance of this particular criterion with a panel of SRI experts representing different stakeholders from the field. This, however, falls beyond the scope of the current paper, but presents an interesting alley for future research.

In addition to the average interval width, we also calculated the average absolute distance between the weights from the first and the second expert in the first and fifth summary column. We find this distance to be 4.97% (2.81% for the MACBETH weights), which is quite significant considering that an equal-weighting of the criteria would lead to individual weights of 5%. The average distance is thus almost 100% of the equal weight. Looking at the differences in more detail, we note several strong deviations, particularly for the first and second criterion. These observations confirm the complexity of the debate on relevant criteria and associated weights to capture social responsibility. It is therefore doubtful whether large SRI expert panels would lead to a useful compromise. Nevertheless, despite this apparent lack in consensus between the two experts, the top of the ranking does not change between the two sets of weights. This confirms earlier research finding that the weights of criteria do not matter all that much. Consequently, we can validate our approach, involving two independent SRI experts and an extensive sensitivity analysis.

From the robustness analysis, we learn that AHP underperforms other eligible MCDA methods. This observation is important, given that earlier research on social responsibility indicators has only implemented the AHP methodology (Pérez-Gladish & M'Zali, 2010; De Moor et al., 2012). Given that three methods prove to be equally robust, we need to consider some additional criteria to determine what method is best for calculating process-oriented social responsibility scores in future research. For the academic application, we first take into account ease of implementation. In order for the method to be feasible for future research in financial economics, it needs to be easy to implement using either existing software packages or by simple programming in spreadsheet or statistical software. Ideally, the software should be available for free to academia. MACBETH is only available upon purchasing the associated MACBETH-M software, which complicates implementation in future research. Some good PROMETHEE software packages are available for free (e.g. Visual PROMETHEE), although we used a costly alternative called D-Sight. With TOPSIS, there is no need for software, as the method can easily be programmed into a spreadsheet program like Microsoft Excel or a statistical software package like R. We also take into account the ease of understanding and transparency of the different methods.

When implementing the obtained scores in mutual fund performance research, it is desirable that it is very clear how the scores were calculated. A conceptually transparent and intuitive methodology is therefore preferred. We consider both TOPSIS and PROMETHEE to be sufficiently transparent and easy to understand. The MACBETH methodology is more complicated and black box and thus less suited for implementation in further financial research. Taking the proposed criteria together, both PROMETHEE and TOPSIS are eligible for implementation in future research on the relationship between social responsibility and financial performance. Note, however, that apart from this particular application of MCDA methodologies, PROMETHEE is considered more advanced and less prone to methodological issues under a variety of data characteristics. TOPSIS is known to sometimes lead to inconsistent results, especially when an alternative has extreme performances on different criteria. Even though such TOPSIS methodological issue was not found in our specific application, we recommend the use of PROMETHEE in future academic applications. PROMETHEE proves to be highly robust, easy to implement and understand, and can also be considered least prone to issues under a variety of possible performance characteristics.

Another application of the scores can be found in a more professional context within the asset management and fund industry. For fund managers, it would be interesting to have a tool that provides the possibility to gain more insights into how scores were calculated and how the design of a new fund could be optimized with respect to social responsibility. For clients, it would be interesting to have a tool that helps them better understand the differences between SRI funds in order to make better investment decisions. Practically, such insights can be obtained by means of a thorough sensitivity analysis. To determine the most recommended methodology for professional applications, we therefore take into account the extensiveness of the available sensitivity analysis tools. Again, we find PROMETHEE to be the most appropriate methodology from the set of robust options. As explained earlier, PROMETHEE can be used in combination with a visual interactive method called GAIA, which greatly adds insights into how a score was obtained and what the best actions would be to improve it. Of most interest is the so-called GAIA plane, which provides insights into the importance of the different criteria, irrespective of the allocated weights. For example, it could very well be that a highly weighted criterion ends up being relatively unimportant in the calculation of the score and the construction of the ranking. Alternatively, low-weighted criteria might be relatively important in the eventual analysis. These kinds of insights can help a fund manager to efficiently detect opportunities for optimizing the design of a fund with respect to social responsibility. Both D-Sight and Visual PROMETHEE provide the software to perform such analyses.

The M-MACBETH software also comes with a built-in sensitivity procedure, although it is less straightforward to determine stability intervals and to gain insights into how the design of a fund can be improved. The MACBETH sensitivity tools are also less graphical, which makes the sensitivity analysis less intuitive for fund managers to use. As TOPSIS is simply programmed into a spreadsheet or statistical software, there is no standard or graphical way of performing sensitivity analyses.

5.5 Application on the Belgian SRI fund market

As we explained before, our small sample of Belgian SRI funds should be considered a first test case for this indicator. As more initiatives will be released to encourage SRI funds to report on their SRI process, it will become easier to access SRI fund information and to include more funds from different countries. For now, we can apply the PROMETHEE indicator on a larger matched-pair sample of funds available on the Belgian market (24 SRI vs. 24 conventional funds). We start from all 24 registered SRI equity funds from the Febelfin sustainable products list and look for 24 matching conventional funds based on six characteristics in the Morningstar database: fund age, fund size, fund type (i.e. accumulation or distribution of gains), geographical orientation, capitalization and investment style. Computing PROMETHEE scores for this sample of 48 funds, we find the following (Table 5.7).

Like for the smaller sample, we find the obtained ranking of funds to be highly in line with generally accepted intuition among practitioners. Also, we note that the 24 SRI funds occupy the top half of the ranking; conventional funds are found at the bottom of the ranking, which makes very good sense. Thematic funds are also ranked lower than SRI funds that apply a broader approach to sustainability.

As indicated before, these scores can have multiple applications. A fund investor can use the scores as an input into the fund selection decision problem. Note that the choice of criteria and weights can be tailored to individual preferences, hence allowing investors to better balance financial and extra-financial criteria in picking mutual funds. Fund managers can use the scores as a way to benchmark other mutual funds on process-level social responsibility. GAIA, the PROMETHEE sensitivity tool, can help fund managers understand how their score can be improved in the most efficient way possible. The scores also accommodate further academic research into the performance differential between SRI and conventional funds. Rather than using the definition from the fund manager on whether a fund is SRI or not, the scores allow for a better approach to study the impact of more or less process-oriented sustainability. The scores can also be used as an input into MCDA sorting tools (e.g. FLOWSORT) to help define different groups of funds based on the process-oriented SRI criteria.

#	Fund	Type	Score	#	Fund	Type	Score
1	23 - Triodos Sustainable Equity Fund	SRI	73.42	25	28 - BNPP Model 6 Classic	Conventional	41.60
2	24 - Triodos Sustainable Pioneer Fund	SRI	73.42	26	27 - KBC New Shares	Conventional	35.04
3	11 - KBC Agri	SRI	69.80	27	30 - KBC Index United States	Conventional	35.04
4	12 - KBC Alternative Energy	SRI	69.80	28	33 - KBC Buyback Europe	Conventional	35.04
5	13 - KBC Climate Change	SRI	69.80	29	34 - KBC Global Leaders	Conventional	35.04
6	14 - KBC Sustainable Euroland	SRI	69.80	30	38 - KBC European Equity	Conventional	35.04
7	15 - KBC Water	SRI	69.80	31	26 - Dexia Europe Innovation	Conventional	33.94
8	16 - KBC World	SRI	69.80	32	40 - Axa Rosenberg	Conventional	33.94
9	2 - Dexia Sustainable EMU	SRI	69.55	33	45 - HSBC European Equity	Conventional	33.94
10	4 - Dexia Sustainable World	SRI	69.55	34	46 - Dexia Europe Classic	Conventional	33.94
11	5 - Dexia Sustainable Europe	SRI	69.55	35	32 - Fidelity World Fund	Conventional	30.07
12	6 - Dexia Sustainable North America	SRI	69.55	36	42 - SSgA World Index Equity	Conventional	30.07
13	7 - Dexia Sustainable Pacific	SRI	69.55	37	36 - Franklin Small-Mid Cap Growth	Conventional	29.81
14	8 - Dexia Sustainable World	SRI	69.55	38	47 - Franklin Global Growth	Conventional	29.81
15	9 - IN.flanders Index Fund	SRI	69.55	39	35 - Transparant B Equity	Conventional	29.66
16	17 - KBC SRI Euro Equities	SRI	69.55	40	29 - Pictet European Equity Selection	Conventional	28.36
17	18 - KBC SRI World Equity	SRI	69.55	41	31 - DWS Top 50 Asia	Conventional	28.36
18	21 - Parvest Sustainable Equity Europe	SRI	69.55	42	37 - R Opal Biens Reels	Conventional	28.36
19	22 - Petercam Equities Europe Sustainable	SRI	67.84	43	41 - Pictet Europe Index	Conventional	28.36
20	10 - ING Sustainable Equity	SRI	65.83	44	43 - Pictet Security P	Conventional	28.36
21	1 - BNPP World Aqua	SRI	65.45	45	25 - Legg Mason Batterymarch	Conventional	25.79
22	19 - Parvest Environmental Opportunities	SRI	65.45	46	39 - Vector Navigator C1	Conventional	25.79
23	20 - Parvest Global Environment	SRI	65.45	47	44 - GAM Star Global Equity Inflation	Conventional	25.79
24	3 - Dexia Sustainable Green Planet	SRI	61.96	48	48 - Universal Invest Quality Growth	Conventional	25.79

Ranking of the matched-pair sample of 48 funds based on PROMETHEE scores. The left panel shows the top-24 funds; the right panel shows the bottom-24 funds.

Table 5.7: *PROMETHEE scores on a larger sample*

Consequently, performance differentials can be tested between the different groups, to further derive insights into the impact of social responsibility on financial performance.

5.6 Conclusion

In our paper we develop a process-oriented social responsibility indicator for mutual funds using multi-criteria decision analysis. This indicator is intended for further application in both academic and professional settings. Upon comparing the range of possible MCDA methods, we find AHP, PROMETHEE, MACBETH and TOPSIS to be eligible for our problem. Twenty criteria are established from the literature, and are combined with an illustrating sample of nine mutual funds available on the Belgian market. Consulting two experts based on an AHP-type questionnaire, we establish relative criteria priorities. We design the performance table in such a way that it can be completed using only publicly disclosed fund information, without any further need for expert judgment. Using the four different MCDA methods, we transform the performance and the weight table into scores and a ranking, which provides more insight into the social responsibility of a mutual fund from the process and design perspective.

The obtained ranking is perfectly consistent across the four considered methods and is also in line with generally accepted intuition in the industry. From the stability intervals and the weights from a second expert, we find the results for PROMETHEE, MACBETH and TOPSIS to be highly robust. We consider two possible applications of the social responsibility scores. First, the scores can be used in academic research to gain more nuanced insights into the performance of SRI funds. Both parametric (e.g. asset pricing; Bauer, Koedijk, & Otten, 2005) and non-parametric (e.g. DEA; Basso & Funari, 2003) methods can draw from the scores to control for differences in process-oriented social responsibility to gain a better understanding of the relationship between return, risk and sustainability. Given its robustness, the ease to both understand and implement the method and its performance under a wide variety of data characteristics, we find PROMETHEE to be the most recommended methodology for academic applications. Second, fund managers can use the MCDA indicator and the scores in a professional setting to optimize the design of newly launched mutual funds with respect to social responsibility and to assist clients in better picking SRI funds. Again, PROMETHEE seems most appropriate, given its extensive range of built-in possibilities to assess sensitivity of obtained results, including the GAIA visual interactive method.

Our paper presents an innovative approach to the design of a process-oriented social responsibility indicator for mutual funds, and determines that PROMETHEE is the recommended methodology for both future academic and professional applications. We contribute to the existing literature by considering the entire set of established MCDA methods, instead of only AHP (Pérez-Gladish & M'Zali, 2010; De Moor et al., 2012) and by determining what the best methodology is for both future academic and professional applications. Additionally, we design the indicator in such a way that it is feasible to implement in future academic and professional applications, without need for elaborate expert or individual judgment, which is in contrast to the different attempts to develop a SRI framework for individual investors (e.g. Hallerbach et al., 2004; Cabello et al., 2014). Future research could have a closer look at the consistent family of criteria and possible cross-cultural differences. Starting from our research, the PROMETHEE-based indicator can now be implemented in SRI performance research to help find out whether top-ranked stocks outperform bottom-ranked stocks. The indicator can also be used in professional applications to optimize the design of new mutual funds with respect to social responsibility and to help clients select SRI funds. In addition to our illustration for the Belgian market, the proposed indicator could also be applied to other markets. Finally, it could be insightful to develop a similar MCDA indicator that yields classes of funds with respect to social responsibility, rather than quantitative scores (e.g. using the PROMETHEE-based FLOWSORT methodology). Such a tool could, for example, be helpful for regulators to help classify and label mutual funds.

Appendix 5.A: Description of criteria

- Priority screening process: the fund first executes the screening process, after which a financial analysis is implemented (not the other way around);
- Independent data gathering and analysis of sustainability:
  - An independent external specialist company (e.g. EIRIS) gathers the necessary data and analyzes sustainability;
  - SRI principles established by national (e.g. Febelfin) and international (e.g. United Nations) organizations are referred to and reflected in the portfolio selection criteria;
  - NGOs and relevant stakeholders are involved in the data gathering process.
- Positive selection criteria;
  - A best-in-class approach (top 30% performing companies in an industry) is developed with respect to ESG criteria;
  - Sector specific criteria are used;
  - Investment is principally ($> 75\%$) in companies that actively invest in sustainable technologies (e.g. green electricity, CO_2 reducing machinery, waste reduction, water quality).
- Negative selection criteria;
  - Categorical rejects using predefined exclusion criteria (e.g. companies involved with nuclear power, tobacco and/or weapons);
  - Contestable activities (e.g. gambling, genetically modified organisms, bio hazards) can lead to exclusion, depending on the

extent of involvement and the context (more nuanced than categorical rejects);

- No investments in unsustainable technologies, irrespective of possible social or ecological measures.

    * Unsustainable technologies: coal plants, nuclear energy, crude oil, coal to liquid, macro-scale hydro power.

- Monitoring and updating;
    - A research team checks legal and regulatory developments, trends and behavior of companies such that criteria remain in line with recent societal developments;
    - Portfolio is monitored for compliance with the set of defined criteria (continuously, sector or event specific).
- Dialogue;
    - Companies are informed about conclusions of the fund's research and get suggestions for improvement of social performance;
    - There is an active engagement policy, which means that there is a constructive and critical dialogue with the companies included in the fund's portfolio in light of positively influencing corporate behavior;
    - There is an active voting policy, which means that representatives of the fund attend shareholder meetings, speak up and vote to change companies' behavior for the better.
- Transparency;
    - Release of qualitative information about the screening process (e.g. criteria used, description of process);

  - Release of quantitative information about the screening process (e.g. scores for individual funds, investment universe);
  - The composition of the portfolio is continuously disclosed;
  - The fund complies with national and international transparency guidelines (e.g. Eurosif/Belsif).

- Board of experts: a board of experts is consulted to help develop the methodology for building the portfolio.

Appendix 5.B: Performance tables

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20
BNPP S	1	1	1	0	0	0	1	1	1	0	1	1	0	1	1	1	0	1	1	1
Dexia S	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	0	1	1	1
ING S	1	1	1	1	1	0	0	1	1	0	1	1	1	1	1	1	0	1	1	1
KBC S	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1
Triodos	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1
BNPP C	0	1	1	0	0	0	0	1	1	0	1	0	0	1	1	0	0	1	0	0
Dexia C	0	0	1	0	0	0	0	1	0	0	0	0	0	1	1	0	0	1	0	0
ING C	0	0	1	0	0	0	0	1	1	0	1	0	0	1	1	0	0	1	0	0
KBC C	0	1	1	0	0	0	0	1	0	0	0	0	0	1	1	0	0	1	0	0

Scores for the 9 funds in our initial sample with respect to the 20 process-level SRI criteria.

BNPP S	BNPP L1 Equity World Aqua
Dexia S	Dexia L Sustainable World
ING S	ING L Invest Sustainable Equity
KBC S	KBC EcoFund World
Triodos	Triodos Sustainable Pioneer Fund
BNPP C	BNPP L1 Equity World
Dexia C	Dexia Quant Equities World
ING C	ING L Invest World
KBC C	KBC Equity Fund World

Explanation of fund abbreviations.

C1	Priority screening process
C2	Data gathering and analysis of sustainability by independent external specialists (e.g. EIRIS)
C3	Incorporation of SRI principles established by reputable organizations (e.g. UN SRI, Febelfin)
C4	Information from stakeholders and relevant NGOs
C5	Best-in-class approach with respect to ESG criteria
C6	Use of sector specific positive criteria
C7	Investment is principally (>75%) in companies that invest in sustainable technologies
C8	Use of categorical rejects
C9	Assessment by means of negative criteria
C10	Exclusion of unsustainable technologies
C11	A research team checks legal and regulatory developments, trends and behavior of companies such that criteria are in line with recent societal developments
C12	Monitoring if portfolio is consistent with defined criteria (continuously, sector specific or occasion specific)
C13	Companies are informed about conclusions selection methodology
C14	Active engagement policy (constructive and critical dialogue with companies in portfolio)
C15	Active voting policy (voting at companies' shareholder meetings)
C16	Release of qualitative information about the screening process (e.g. applied screens)
C17	Release of quantitative information about the screening process (e.g. scores)
C18	Release of current portfolio
C19	Compliance with external transparency guidelines (e.g. Eurosif/Belsif)
C20	Board of experts

Explanation of criteria abbreviations.

Appendix 5.C: AHP questionnaire

Please fill out the following tables. The purpose is to use pairwise comparisons to find the weight of the different criteria (Table 5.2) within their group. In comparing two criteria, you need to state which criterion you deem most important (priority), and how much more important you deem it (intensity). The intensity scale goes from 1 to 9.

Intensity scale pairwise comparison		
Intensity	*Definition*	*Explanation*
1	Equally important	The two criteria equally contribute to the goal.
3	Slightly more important	One criterion is slightly more contributing to the goal.
5	More important	One criterion is clearly more important with respect to the goal.
7	Much more important	One criterion dominates the other in light of the goal.
9	Extremely more important	One criterion is most definitely more important with respect to the goal.
Intensities 2, 4, 6 and 8 can be used to indicate intermediary values. Intensities, 1.1, 1.2, 1.3 etc. can be used to indicate even more nuance.		

Pairwise comparison of criteria

Criterion		**Priority**	**Intensity**
A	*B*	*A or B*	*(1 - 9)*
Screening process and consistency	Transparency and control		

Pairwise comparison of subcriteria

Transparency, reporting and control criterion		**Priority**	**Intensity**
A	*B*	*A or B*	*(1 - 9)*
Transparency	Board of experts		

Screening and process consistency criterion		Priority	Intensity
A	*B*	*A or B*	*(1 - 9)*
Priority screening process	Independent data gathering and analysis of sustainability		
Priority screening process	Positive selection criteria		
Priority screening process	Negative selection criteria		
Priority screening process	Monitoring and updates		
Priority screening process	Dialogue		
Independent data gathering and analysis of sustainability	Positive selection criteria		
Independent data gathering and analysis of sustainability	Negative selection criteria		
Independent data gathering and analysis of sustainability	Monitoring and updates		
Independent data gathering and analysis of sustainability	Dialogue		
Positive selection criteria	Negative selection criteria		
Positive selection criteria	Monitoring and updates		
Positive selection criteria	Dialogue		
Negative selection criteria	Monitoring and updates		
Negative selection criteria	Dialogue		
Monitoring and updates	Dialogue		

Pairwise comparison of sub-subcriteria

Independent data gathering and analysis of sustainability subcriterion		Priority	Intensity
A	*B*	*A or B*	*(1 - 9)*
Data gathering and analysis of sustainability by independent external specialists	Incorporation of SRI principles established by reputable organizations		
Data gathering and analysis of sustainability by independent external specialists	Information from stakeholders and relevant NGOs		
Incorporation of SRI principles established by reputable organizations	Information from stakeholders and relevant NGOs		

Positive selection subcriterion		Priority	Intensity
A	*B*	*A or B*	*(1 - 9)*
Best-in-class approach for criteria with respect to ESG criteria	Use of sector specific positive criteria		
Best-in-class approach for criteria with respect to ESG criteria	Investment is principally (>75%) in companies that invest in sustainable technologies		
Use of sector specific positive criteria	Investment is principally (>75%) in companies that invest in sustainable technologies		

Negative selection subcriterion		Priority	Intensity
A	*B*	*A or B*	*(1 - 9)*
Use of categorical rejects	Assessment by means of negative criteria		
Use of categorical rejects	Exclusion of unsustainable technologies		
Assessment by means of negative criteria	Exclusion of unsustainable technologies		

Monitoring and updates subcriterion		Priority	Intensity
A	*B*	*A or B*	*(1 - 9)*
A research teams checks legal and regulatory developments, trends and behavior of companies such that criteria are in line with recent societal developments	Monitoring if portfolio is consistent with defined criteria (continuously, sector specific or occasion specific)		

Dialogue subcriterion		Priority	Intensity
A	*B*	*A or B*	*(1 - 9)*
Companies are informed about conclusions selection methodology	Active engagement policy (constructive and critical dialogue with companies in portfolio)		
Companies are informed about conclusions selection methodology	Active voting policy (voting at companies' shareholder meetings)		
Active engagement policy (constructive and critical dialogue with companies in portfolio)	Active voting policy (voting at companies' shareholder meetings)		

Transparency subcriterion		Priority	Intensity
A	*B*	*A or B*	*(1 - 9)*
Release of qualitative information about the screening process (e.g. applied screens)	Release of quantitative information about the screening process (e.g. scores)		
Release of qualitative information about the screening process (e.g. applied screens)	Release of current portfolio		
Release of qualitative information about the screening process (e.g. applied screens)	Compliance with external transparency guidelines		
Release of quantitative information about the screening process (e.g. scores)	Release of current portfolio		
Release of quantitative information about the screening process (e.g. scores)	Compliance with external transparency guidelines		
Release of current portfolio	Compliance with external transparency guidelines		

Appendix 5.D: More details on MCDA methods

In this appendix we revisit the five implemented MCDA strategies in some more detail (cfr. Ishizaka & Nemery, 2013; Belton & Stewart, 2002; Figueira et. al, 2005).

As noted before, every MCDA technique consists of five essential steps:

(i) Establish assessment criteria;

(ii) Pick decision alternatives;

(iii) Score alternatives with respect to every criterion;

(iv) Weigh every criterion;

(v) Aggregate information into a final score.

The specificity of every technique can be in different parts of this process.

Analytic hierarchy process

The specificity of the analytic hierarchy process (AHP) is in the elicitation of weights, i.e. step four. The performance matrix is simply formed from step one through three, after which the relative importance of every criterion is established using pairwise comparisons. An example of these pairwise comparisons can be found in Appendix 5.C.

In the AHP pairwise comparisons are made using ratio scales, i.e. by relative judgement of two criteria using an intensity scale going from one to nine (another scale could also be feasible). From the relative judgments, comparison matrices can be constructed.

An important assumption before deriving priorities from the ratio scale comparison matrix, is that judgments are consistent. To test this assumption, Saaty (1977) developed the so-called consistency index (CI):

$$CI = \frac{\lambda_{max} - n}{n - 1}$$

Where n is the number of criteria and λ_{max} is the maximum eigenvalue. From the consistency index, the consistency ratio (CR) can be calculated:

$$CR = \frac{CI}{RI}$$

With RI being the CI of 500 randomly filled matrices. If the value of CR is less than 10%, i.e. the inconsistency level is less than 10% of 500 randomly filled matrices, the comparison matrix is acceptably consistent.

Once we have a consistent comparison matrix, priorities can be derived. The vector of priorities **p** can be found solving the following equation:

$$\mathbf{A}\mathbf{p} = m\mathbf{p}$$

Where **A** is the consistent comparison matrix and m is the dimension of **A**. The solution to this simple equation yields the criteria weights, which can then be used in a simple weighted sum to find the eventual AHP scores. Note that most AHP software performs calculations on a non-normalized scale coming from the underlying ratio judgements, which might lead to very small scoring differentials when normalizing to a scale between 0 and 100.

Multi-attribute utility theory

The special feature of multi-attribute utility theory (MAUT) is in the preference elicitation and transformation of scores. Typically, the performance matrix is transformed to a utility table using preference functions that are defined for each and every criterion. The rest of the MAUT process is simply a weighted sum using the transformed performance matrix.

As explained in this chapter, our choice for binary-coded criteria leads to simple discrete utility functions, causing the performance matrix to be indifferent from the utility matrix. As we are also using weights elicited under the AHP, there is no difference between the AHP and MAUT results in our example.

MACBETH

Next to AHP and MAUT, MACBETH (Measuring Attractiveness by a Categorical Based Evaluation Technique) is yet another value measurement method. Just like AHP, MACBETH builds from pairwise comparisons, but using an interval scale rather than a ratio scale, i.e. the decision maker needs to provide judgment on the difference of attractiveness between alternatives.

The process of MACBETH is very similar to AHP. From the hierarchy of criteria, we score the different alternatives with respect to every criterion inside the performance matrix. The matrix of judgements is simply adopted from the AHP judgments in our case, to circumvent an excessive amount of pairwise comparisons (see Section 5.2.4 for a further motivation). Since the AHP judgments are already tested for consistency, the next step is to transform the matrix of judgments to a weighting matrix using the MACBETH linear program (LP). This LP is what differentiates results from MACBETH with AHP.

The objective function of the LP is (Bana e Costa & Vansnick, 1994):

$$minimize\ \Phi(o_1)$$

With $\Phi(o_1)$ the score of the most attractive alternative o_1. Note that maximization would lead to infinite solutions. The decision variables are the scores of all alternatives:

$$\Phi(o_i), i \in \{1, 2, \ldots, n\}$$

There are three groups of constraints:

(i) **Ordinal constraints**

- $\forall o_i, o_j, i, j \in \{1, 2, \ldots, n\}$: o_i is preferred to $o_j \Rightarrow \Phi(o_i) \geq \Phi(o_j) + \delta(i, j)$
  - $\delta(i, j)$ is the difference of attractiveness between o_i and o_j.
- $\forall o_i, o_j, i, j \in \{1, 2, \ldots, n\}$: o_i and o_j are indifferent $\Rightarrow \Phi(o_i) = \Phi(o_j)$

(ii) **Semantic conditions**

- $\forall o_i, o_j, o_k, o_l, i, j, k, l \in \{1, 2, \ldots, n\}$: $\Phi(o_i) - \Phi(o_j) \geq \Phi(o_k) - \Phi(o_l) + \delta(i, j, k, l)$
  - $\delta(i, j, k, l)$ is the number of semantic categories between the difference of attractiveness between o_i and o_j, and the difference of attractiveness between o_k and o_l.

(iii) **Grounding conditions**

- $\Phi(o_n) = 0$
  - Score of the least attractive alternative is zero.

The problem with this LP and MACBETH in general is that the solution is not always unique. In our case, however, we do find satisfactory results. Using the weights on the MACBETH scale, we can again perform a simple weighted sum operation to find the MACBETH scores.

PROMETHEE

An alternative to the value measurement approach can be found in an outranking technique like PROMETHEE (Preference Ranking Organization

METHod for Enriched Evaluation). Remember that we opt for binary scored criteria and a discrete preference function to facilitate the process and further implementation (see Section 5.2.4).

From the performance matrix, we start by constructing unicriterion preference degrees. This means that we have 20 unicriterion preference tables with 9 x 9 elements (number of alternatives). On the diagonal, we have all 0's as an alternative can not be preferred to itself. The other elements of the preference tables are either 1 or 0, depending on whether there is a difference in score (1; i.e. preference) or not (0; i.e. indifference).

In a next step, the unicriterion preference degrees are summarized in unicriterion positive, negative and net flows. The unicriterion positive flow is simply the normalized sum of each row, excluding the diagonal elements (i.e. how is an alternative preferred to the other alternatives, on average). The unicriterion negative flow is the normalized sum of each column, excluding the same diagonal elements (i.e. how are other alternatives preferred over an alternative, on average). Subtracting the negative from the positive unicriterion flows yields a unicriterion net flow for each alternative.

A global positive, negative and net flow, indicating how an alternative is globally preferred to other alternatives, follows from a weighted sum using the AHP weights. The PROMETHEE II ranking that we use in our analysis draws from the global net flows and leads to a complete ranking of alternatives.

TOPSIS

The technique for order preference similarity to the ideal solution or TOPSIS comes from the goal, aspiration and reference-level category of MCDA methods. The principle of the technique is very simple: define an ideal and anti-ideal solution and calculate the ratio of distances from every alternative to these two theoretically defined solutions.

The first steps are again very similar to the other MCDA techniques. We start from the performance matrix and weigh the scores using weights coming from the AHP. Next, we calculate the distance from every alternative to the ideal (matrix full of 1's; v_i^+) and anti-ideal solution (matrix full of 0's; v_i^-), i.e. the positive (d_a^+) and negative distance (d_a^-).

$$d_a^+ = \sqrt{\sum_i (v_i^+ - v_{ai})^2}, \quad a = 1, \ldots, m$$

$$d_a^- = \sqrt{\sum_i (v_i^- - v_{ai})^2}, \quad a = 1, \ldots, m$$

Finally, we calculate the relative closeness coefficient of every alternative:

$$C_a = \frac{d_a^-}{d_a^+ + d_a^-}$$

Normalizing these C_a's, we again obtain a score between 0 and 100 for every alternative.

Chapter 6

Sorting mutual funds with respect to process-oriented social responsibility: A FLOWSORT application

Abstract

We establish a robust FLOWSORT-based tool to sort mutual funds with respect to process-oriented social responsibility and recommend the use of limiting profiles with open classes. The tool provides an alternative for the limited dichotomous classification of funds, i.e. socially responsible investing (SRI) versus conventional funds. By allowing for more heterogeneity in social responsibility the sorting tool is promising for scholars to improve fund performance measurements, and useful for governments to better regulate the supply of SRI products. We also implement the tool in a case study for the Belgian SRI market.

6.1 Introduction

Over the course of the last decade, socially responsible investing (SRI) has become a mainstream investment strategy. Instead of only considering financial objectives, many investors now take into account environmental, social and governance issues as well. A typical motivation for SRI is trying to do financially well while doing socially good. However, researchers are interested in the question whether SRI makes any financial sense. Implementing multi-factor asset pricing regressions, which take into account several factors of risk, most researchers either find a significant underperformance of SRI funds, or no performance differential at all. The problem with the current approach is that no heterogeneity in terms of social responsibility is taken into account, as risk-adjusted returns from both a sample of SRI and conventional funds are simply tested for statistical significant differences. Hence the investment universe is falsely reduced to SRI vs. non-SRI. For a more comprehensive overview of the literature, we refer to several excellent review papers (e.g. Margolis & Walsh, 2003; Orlitzky, Schmidt, & Rynes, 2003).

A helpful way to circumvent the dichotomous SRI versus conventional fund approach is multi-criteria decision analysis (MCDA). This operations research/decision sciences methodological framework provides the tools to deal with situations that call for simultaneous consideration of multiple conflicting decision factors. Five steps are central to MCDA (Belton & Stewart, 2002): establishing assessment criteria, defining alternatives, scoring alternatives, weighing criteria and aggregating all of this information. MCDA can address four types of "problematiques" (Roy, 1996): picking, sorting, ranking and describing. In this paper, we present a MCDA sorting tool as a way to distinguish funds based on process-oriented social responsibility criteria. A MCDA-based scoring tool has already been presented by Verheyden and De Moor (2014). The benefit of sorting over scoring tools is that the significance of small performance differentials is reflected in the fact whether a fund is sorted into a superior/inferior category or not.

To the best of our knowledge, this paper is the first attempt to build a MCDA-based tool to sort mutual funds with respect to social responsibility. We find the use of limiting profiles with open classes to be most recommended and design the sorting tool in a way that it can be instrumental for implementation in future mutual fund performance research. For example, scholars could apply multi-factor asset pricing regressions to test for significantly different risk-adjusted returns between the five proposed ordered categories, enriching the typical dichotomous distinction between SRI and non-SRI funds. We include a case study for the Belgian SRI market to illustrate this potential use of the sorting tool. The proposed categories could also be used to construct a factor mimicking zero-investment portfolio to control for an "ethics risk factor", following an earlier attempt by Renneboog, Ter Horst and Zhang (2008b). Finally, governments might profit from the sorting tool to help regulate the supply of SRI funds (e.g. government-issued SRI labels).

6.2 Data and methodology

To build the sorting tool we implement the five building blocks of the MCDA framework. The first step involves the establishment of assessment criteria. As we aim to assess social responsibility on the aggregate level of a fund, and not on the individual level of a single stock, we opt for criteria that describe the investment process of a fund in terms of social responsibility, hence we refer to process-oriented social responsibility. Table 6.1 presents our hierarchy of criteria, which was built from earlier research (Pérez-Gladish & M'Zali, 2010; De Moor, Devooght and De Bondt, 2012) and directives on SRI by the United Nations (2013) and Febelfin (2012), the Belgian federation of the financial industry.

Goal	Criteria	Subcriteria	Sub-subcriteria	Weight E1	Weight E2
Social performance indicator	Screening process and consistency	C1. Priority screening process		33.15%	2.40%
		Independent data gathering and analysis of sustainability	C2. Data gathering and analysis of sustainability by independent external specialists (e.g. EIRIS)	2.16%	23.61%
			C3. Incorporation of SRI principles established by reputable organizations (e.g. UN SRI, Febelfin)	5.03%	3.27%
			C4. Information from stakeholders and relevant NGOs	0.56%	6.80%
		Positive selection criteria	C5. Best-in-class approach with respect to ESG criteria	5.74%	9.20%
			C6. Use of sector specific positive criteria	2.76%	2.13%
			C7. Investment is principally (>75%) in companies that invest in sustainable technologies	0.49%	4.43%
		Negative selection criteria	C8. Use of categorical rejects	7.57%	2.47%
			C9. Assessment by means of negative criteria	7.57%	5.14%
			C10. Exclusion of unsustainable technologies	7.57%	3.56%
		Monitoring and updates	C11. A research teams checks legal and regulatory developments, trends and behavior of companies such that criteria are in line with recent societal developments	5.25%	1.97%
			C12. Monitoring if portfolio is consistent with defined criteria (continuously, sector specific or occasion specific)	1.75%	5.91%
		Dialogue	C13. Companies are informed about conclusions selection methodology	0.24%	1.93%
			C14. Active engagement policy (constructive and critical dialogue with companies in portfolio)	0.51%	0.93%
			C15. Active voting policy (voting at companies' shareholder meetings)	2.83%	1.34%
	Transparency and control	Transparency	C16. Release of qualitative information about the screening process (e.g. applied screens)	4.00%	3.90%
			C17. Release of quantitative information about the screening process (e.g. scores)	4.00%	3.90%
			C18. Release of current portfolio	2.23%	9.13%
			C19. Compliance with external transparency guidelines (e.g. Eurosif/Belsif)	0.91%	1.80%
		C20. Board of experts		5.56%	6.25%

The hierarchy of criteria starts from the overall goal of the MCDA analysis, followed by different levels of criteria. The 20 bottom-level criteria are used in the eventual sorting exercise. To allow for robustness checks, we use weights from two independent SRI experts (E1 and E2).

Table 6.1: *Hierarchy of criteria and weights*

In a second step we define a set of alternatives, i.e. mutual funds. We opt for the matched-pair sampling approach, as the implementation of the proposed MCDA methodology calls for transparent information on how a fund was constructed and what criteria were taken into account. First, we establish a sample of registered SRI funds, for which the necessary information tends to be hard to retrieve and unreliable (Tippet, 2001; Hogget & Nahan, 2002). As a global leader in the promotion of SRI products, the Belgian financial sector organization (Febelfin) has constructed a list of sustainable products that includes all registered SRI mutual funds that comply with a minimum recommendation on what is considered socially responsible (Febelfin, 2012). To be included in this list, transparent and detailed information on the construction and policies of the SRI mutual fund needs to be released. Because of the availability of this framework and the associated information, we decide to implement our methodology on the Belgian mutual fund market. Given the globally improving regulatory SRI framework, future research can implement a similar approach on other markets as well.

We start from the 24 registered SRI equity funds that are included in the Febelfin list, and construct a matched-pair sample of 24 conventional mutual funds that are traded on the Belgian market, using the Morningstar mutual fund database. For the matching process we use six fund characteristics that are included in Morningstar: fund age, fund size, fund type (i.e. accumulation or distribution of gains), geographical orientation, capitalization and investment style. For all criteria except fund age, the matching was exact, i.e. the SRI fund and the matched conventional fund have the exact same characteristic. Given the match on the other five criteria, we then selected that conventional fund that was closest to the age of the associated SRI fund. Note that we do not explicitly control for survivorship bias, as we only consider funds that are still trading today. The main reason for this is that we do not have the necessary information on dead funds to construct the MCDA indicator. However, given our extensive matching-pair efforts, we believe this will not interfere with our results. The final sample of 48 funds is listed in Table 6.2, together with the benchmarks used to perform our analysis.

Next, we need to score the alternatives with respect to the 20 criteria. For every alternative, we assess whether the different criteria apply (1) or not (0) using publicly disclosed information (e.g. fund prospectus, website information, transparency documents from the Febelfin website). The reason for using binary assessments for the individual criteria is to enhance the replicability of the sorting tool for future applications in finance, by avoiding the need for elaborate expert judgments. Since we aggregate all of these assessments across the criteria and the alternatives using MCDA techniques, the eventual scores used to build the categories are no longer dichotomous, and thus better reflect heterogeneity. The performance table can be found in Appendix 5.B.

Prior to calculating the scores, we also need to indicate the relative importance of the different criteria. To do so we ask two independent SRI experts to fill out a questionnaire that asks for pairwise comparisons of the different criteria (Appendix 5.C). Asking two independent experts allows us to test for robustness of results. From these comparisons we can calculate weights for the different criteria using the analytic hierarchy process (Saaty, 1980). This is the only step where we allow for expert judgment. The weights are represented in Table 6.1.

In our final step we construct categories using FLOWSORT, which draws from PROMETHEE II rankings to assign alternatives to categories using central and limiting profiles. PROMETHEE is the acronym for "Preference Ranking Organization METHod for Enriched Evaluation" and was originally developed by Brans and Vincke (1985). It belongs to the outranking school of MCDA methods and starts from the notion that "one solution outranks another if it is at least as good as the other in most respects, and not too much worse in any one respect" (Belton & Stewart, 2002). Starting from preference degrees that reflect a decision maker's attitude towards the different criteria, PROMETHEE II constructs a complete ranking comput-

	SRI Alternative	ISIN-code		Conventional alternative	ISIN-code
A1	BNP Paribas L1 Equity World Aqua	LU0831546592	A25	Legg Mason Batterymarch Global Equity Fund A	IE00B5589395
A2	Dexia Equities L Sustainable EMU	LU0344047559	A26	Dexia Equities L Europe Innovation C	LU0344046155
A3	Dexia Equities L Sustainable Green Planet	LU0304860991	A27	KBC Equity Fund - New Shares	BE0170533070
A4	Dexia Equities L Sustainable World	LU0113400328	A28	BNP Paribas L1 Model 6 Classic	LU0377118962
A5	Dexia Sustainable Europe	BE0173540072	A29	Pictet-European Equity Selection-R	LU0130732109
A6	Dexia Sustainable North America	BE0173901779	A30	KBC Index Fund United States	BE0166769266
A7	Dexia Sustainable Pacific	BE0174191768	A31	DWS Invest Top 50 Asia	LU0145648886
A8	Dexia Sustainable World	BE0946893766	A32	Fidelity Funds - World Fund E	LU0115769746
A9	IN.flanders Index Fund	BE0175210286	A33	KBC Equity Fund - Buyback Europe	BE0174407016
A10	ING (L) Invest Sustainable Equity	LU0119216553	A34	KBC Equity Fund - Global Leaders	BE0174807132
A11	KBC Eco Fund Agri	BE6222656090	A35	Transparant B Equity	BE0935007246
A12	KBC Eco Fund Alternative Energy	BE0175280016	A36	Franklin Global Small-Mid Cap Growth	LU0144644332
A13	KBC Eco Fund Climate Change	BE0946844272	A37	R Opal Biens Réels F	FR0010563064
A14	KBC Eco Fund Sustainable Euroland	BE0175718510	A38	KBC Institutional Fund European Equity Classic	BE0176222702
A15	KBC Eco Fund Water	BE0175479063	A39	Vector Navigator C1	LU0172125329
A16	KBC Eco Fund World	BE0133741752	A40	AXA Rosenberg Global Equity Alpha Fund A	IE0008366811
A17	KBC Institutional Fund SRI Euro Equities	BE0175761940	A41	Pictet-Europe Index-R	LU0130731713
A18	KBC Institutional Fund SRI World Equity	BE0168344498	A42	SSgA World Index Equity Fund P	FR0000018277
A19	Parvest Environmental Opportunities	LU0406802339	A43	Pictet-Security-P	LU0270904781
A20	Parvest Global Environment	LU0347711466	A44	GAM Star Global Equity Inflation Focus C	IE00B5BDSJ79
A21	Parvest Sustainable Equity Europe	LU0212189012	A45	HSBC Global Investment Funds European Equity EC	LU0164863887
A22	Petercam Equities Europe Sustainable	BE0940002729	A46	Dexia Quant Equities Europe Classic C	LU0149700378
A23	Triodos Sustainable Equity Fund	LU0278271951	A47	Franklin Global Growth A	LU0122613069
A24	Triodos Sustainable Pioneer Fund	LU0278272843	A48	Universal Invest Quality Growth B	LU0124604223

The list of alternatives consists of 24 SRI and 24 matched conventional funds (incl. the ISIN code) from the Belgian market. Our sorting tool will yield in 5 categories by introducing more heterogeneity between these 2 naïve categories, which are typically used in SRI performance research.

Table 6.2: *List of alternatives*

ing and aggregating unicriterion flows that indicate how one alternative is preferred to another for every single criterion. FLOWSORT, originally developed by Nemery and Lamboray (2008), takes the PROMETHEE II net flow scores to assess the relative position of alternatives with respect to reference profiles and hence assigns the alternatives to completely ordered categories. Two types of reference profiles can be implemented: limiting profiles or central profiles. Limiting profiles define the boundaries between the different categories. We distinguish two options: open and closed categories. On top of the intra-category boundaries, the closed option also requires a boundary on the bottom of the lowest category and a boundary on top of the highest category. That way, alternatives can also be discontinued from any possible category. We choose for open categories, as we want all funds to be assigned to a certain group to account for heterogeneity. Central profiles use representative alternatives for each group, rather than boundaries between groups. An important condition for both types of approaches is that the different categories must dominate each other. We define and implement both open limiting profiles and central profiles, building from expert information and several performance profiles that are drawn from the performance table. From the performance profiles, five categories become apparent and thus four open limiting profiles and five central profiles are established for each expert (cf. Table 6.3). We compare the sorting between both experts to test for robustness.

An important advantage of FLOWSORT over most other sorting techniques (e.g. Zopounidis & Doumpos, 2002; Araz & Ozkarahan, 2007) is that the allocation of an alternative to a group is independent from the allocation of another alternative. In addition we prefer a PROMETHEE-based ranking approach as the PROMETHEE ranking methodology has proven to be superior to other approaches in assessing process-oriented social responsibility of mutual funds (Verheyden & De Moor, 2014).

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20
Expert 1																				
Limiting profiles (open classes)																				
Profile 4	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1
Profile 3	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	0	1	1	1
Profile 2	1	1	1	1	0	0	0	1	1	0	1	1	1	0	0	0	0	1	1	1
Profile 1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0
Central profiles																				
	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20
Profile 5	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1
Profile 4	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	0	1	1	1
Profile 3	1	1	1	1	0	0	0	1	1	0	1	1	1	0	0	0	0	1	1	1
Profile 2	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0
Profile 1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
Expert 2																				
Limiting profiles (open classes)																				
	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20
Profile 4	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1
Profile 3	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	0	1	1	1
Profile 2	1	1	1	1	0	0	0	1	1	0	1	1	1	0	0	0	0	1	1	1
Profile 1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0
Central profiles																				
	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20
Profile 5	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1
Profile 4	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	0	1	1	1
Profile 3	1	1	1	1	0	0	0	1	1	0	1	1	1	0	0	0	0	1	1	1
Profile 2	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0
Profile 1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0

For every expert we define 4 limiting and 5 central profiles, which simply are theoretically defined alternatives with a particular score on the 20 different criteria (C1-C20). The profiles have been established from the preferences expressed by the experts and information from the performance table, which points to 5 distinguished performance profiles.

Table 6.3: *Limiting (open classes) and central profiles*

6.3 Results and discussion

We implement the FLOWSORT method in the Smart Picker Pro software. The ordered sorting of the funds in five categories can be found in Table 6.4. Overall we see quite consistent sorting across the two different types of profiles and the two experts, which adds robustness to the results. Most striking is the perfect consistency in the sorting of the top-tier alternatives, i.e. the SRI funds by Triodos and KBC. Triodos is a niche player in the banking industry that promotes itself as “the sustainable bank.” KBC is a traditional commercial bank, but with a long-standing tradition in SRI and a holistic approach to the design of SRI funds. These result are thus not surprising and in line with generally accepted intuition in the industry.

If we compare the results between the inputs provided by both independent experts, we see some differences. Most notably, the ranking within the top group changes between Triodos and KBC. However, the FLOWSORT method has considered this difference to be insignificant and thus sorted SRI funds from both providers in the top category. This kind of additional interpretation of differences in ranking and scores is exactly the added value of FLOWSORT over the ranking and scoring tools. Besides, we see that the limiting profile sorting remains robust over the two experts; for the central profile there are some mild differences in the sorting of lower-tier funds. Despite the rather large differences in the expert judgments, we see that overall results are fairly robust. In addition to the robustness of the limiting profile across both experts, it is also easier to implement because one less predefined profile is required. Taking into account the implementation of these sorting groups in asset pricing regressions, the limiting profiles are also preferred because they yield more balanced groups, whereas the central profiles lead to a disparity of large and small groups.

Expert 1						Expert 2					
Limiting profile (open classes)			Central profile			Limiting profile (Open classes)			Central profile		
Alternatives	*Group*	*Flow*	*Alternatives*	*Group*	*Flow*	*Alternatives*	*Group*	*Flow*	*Alternatives*	*Group*	*Flow*
A23 - Triodos Sustainable Equity Fund	1	0.29691	A23 - Triodos Sustainable Equity Fund	1	0.42407	A11 - KBC Agri	1	0.26150	A11 - KBC Agri	1	0.37600
A24 - Triodos Sustainable Pioneer Fund	1	0.29691	A24 - Triodos Sustainable Pioneer Fund	1	0.42407	A12 - KBC Alternative Energy	1	0.26150	A12 - KBC Alternative Energy	1	0.37600
A11 - KBC Agri	1	0.22602	A11 - KBC Agri	1	0.35318	A13 - KBC Climate Change	1	0.26150	A13 - KBC Climate Change	1	0.37600
A12 - KBC Alternative Energy	1	0.22602	A12 - KBC Alternative Energy	1	0.35318	A14 - KBC Sustainable Euroland	1	0.26150	A14 - KBC Sustainable Euroland	1	0.37600
A13 - KBC Climate Change	1	0.22602	A13 - KBC Climate Change	1	0.35318	A15 - KBC Water	1	0.26150	A15 - KBC Water	1	0.37600
A14 - KBC Sustainable Euroland	1	0.22602	A14 - KBC Sustainable Euroland	1	0.35318	A16 - KBC World	1	0.26150	A16 - KBC World	1	0.37600
A15 - KBC Water	1	0.22602	A15 - KBC Water	1	0.35318	A23 - Triodos Sustainable Equity Fund	1	0.25350	A23 - Triodos Sustainable Equity Fund	1	0.36800
A16 - KBC World	1	0.22602	A16 - KBC World	1	0.35318	A24 - Triodos Sustainable Pioneer Fund	1	0.25350	A24 - Triodos Sustainable Pioneer Fund	1	0.36800
A2 - Dexia Sustainable EMU	2	0.22112	A2 - Dexia Sustainable EMU	2	0.34828	A2 - Dexia Sustainable EMU	2	0.21750	A2 - Dexia Sustainable EMU	2	0.33200
A4 - Dexia Sustainable World	2	0.22112	A4 - Dexia Sustainable World	2	0.34828	A4 - Dexia Sustainable World	2	0.21750	A4 - Dexia Sustainable World	2	0.33200
A5 - Dexia Sustainable Europe	2	0.22112	A5 - Dexia Sustainable Europe	2	0.34828	A5 - Dexia Sustainable Europe	2	0.21750	A5 - Dexia Sustainable Europe	2	0.33200
A6 - Dexia Sustainable North America	2	0.22112	A6 - Dexia Sustainable North America	2	0.34828	A6 - Dexia Sustainable North America	2	0.21750	A6 - Dexia Sustainable North America	2	0.33200
A7 - Dexia Sustainable Pacific	2	0.22112	A7 - Dexia Sustainable Pacific	2	0.34828	A7 - Dexia Sustainable Pacific	2	0.21750	A7 - Dexia Sustainable Pacific	2	0.33200
A8 - Dexia Sustainable World	2	0.22112	A8 - Dexia Sustainable World	2	0.34828	A8 - Dexia Sustainable World	2	0.21750	A8 - Dexia Sustainable World	2	0.33200
A9 - IN.flanders Index Fund	2	0.22112	A9 - IN.flanders Index Fund	2	0.34828	A9 - IN.flanders Index Fund	2	0.21750	A9 - IN.flanders Index Fund	2	0.33200
A17 - KBC SRI Euro Equities	2	0.22112	A17 - KBC SRI Euro Equities	2	0.34828	A17 - KBC SRI Euro Equities	2	0.21750	A17 - KBC SRI Euro Equities	2	0.33200
A18 - KBC SRI World Equity	2	0.22112	A18 - KBC SRI World Equity	2	0.34828	A18 - KBC SRI World Equity	2	0.21750	A18 - KBC SRI World Equity	2	0.33200
A21 - Parvest Sustainable Equity Europe	2	0.22112	A21 - Parvest Sustainable Equity Europe	2	0.34828	A21 - Parvest Sustainable Equity Europe	2	0.21750	A21 - Parvest Sustainable Equity Europe	2	0.33200
A22 - Petercam Equities Europe Sustainable	3	0.18768	A22 - Petercam Equities Europe Sustainable	2	0.31484	A22 - Petercam Equities Europe Sustainable	3	0.19550	A22 - Petercam Equities Europe Sustainable	2	0.31000
A10 - ING Sustainable Equity	3	0.14833	A10 - ING Sustainable Equity	2	0.27549	A1 - BNPP World Aqua	3	0.14850	A1 - BNPP World Aqua	2	0.26300
A1 - BNPP World Aqua	3	0.14092	A1 - BNPP World Aqua	3	0.26808	A10 - ING Sustainable Equity	3	0.14850	A10 - ING Sustainable Equity	2	0.26300
A19 - Parvest Environmental Opportunities	3	0.14092	A19 - Parvest Environmental Opportunities	3	0.26808	A19 - Parvest Environmental Opportunities	3	0.14850	A19 - Parvest Environmental Opportunities	2	0.26300
A20 - Parvest Global Environment	3	0.14092	A20 - Parvest Global Environment	3	0.26808	A20 - Parvest Global Environment	3	0.14850	A20 - Parvest Global Environment	2	0.26300
A3 - Dexia Sustainable Green Planet	3	0.07254	A3 - Dexia Sustainable Green Planet	3	0.19970	A3 - Dexia Sustainable Green Planet	3	0.09650	A3 - Dexia Sustainable Green Planet	3	0.21100
A28 - BNPP Model 6 Classic	4	-0.32624	A28 - BNPP Model 6 Classic	4	-0.19908	A28 - BNPP Model 6 Classic	4	-0.18550	A28 - BNPP Model 6 Classic	3	-0.07100
A27 - KBC New Shares	4	-0.45460	A27 - KBC New Shares	4	-0.32743	A27 - KBC New Shares	4	-0.25650	A27 - KBC New Shares	4	-0.14200
A30 - KBC Index United States	4	-0.45460	A30 - KBC Index United States	4	-0.32743	A30 - KBC Index United States	4	-0.25650	A30 - KBC Index United States	4	-0.14200
A33 - KBC Buyback Europe	4	-0.45460	A33 - KBC Buyback Europe	4	-0.32743	A33 - KBC Buyback Europe	4	-0.25650	A33 - KBC Buyback Europe	4	-0.14200
A34 - KBC Global Leaders	4	-0.45460	A34 - KBC Global Leaders	4	-0.32743	A34 - KBC Global Leaders	4	-0.25650	A34 - KBC Global Leaders	4	-0.14200
A38 - KBC European Equity	4	-0.45460	A38 - KBC European Equity	4	-0.32743	A38 - KBC European Equity	4	-0.25650	A38 - KBC European Equity	4	-0.14200
A26 - Dexia Europe Innovation	4	-0.47622	A26 - Dexia Europe Innovation	4	-0.34906	A26 - Dexia Europe Innovation	4	-0.49250	A26 - Dexia Europe Innovation	4	-0.37800
A40 - Axa Rosenberg	4	-0.47622	A40 - Axa Rosenberg	4	-0.34906	A40 - Axa Rosenberg	4	-0.49250	A40 - Axa Rosenberg	4	-0.37800
A45 - HSBC European Equity	4	-0.47622	A45 - HSBC European Equity	4	-0.34906	A45 - HSBC European Equity	4	-0.49250	A45 - HSBC European Equity	4	-0.37800
A46 - Dexia Europe Classic	4	-0.47622	A46 - Dexia Europe Classic	4	-0.34906	A46 - Dexia Europe Classic	4	-0.49250	A46 - Dexia Europe Classic	4	-0.37800
A32 - Fidelity World Fund	5	-0.55201	A32 - Fidelity World Fund	4	-0.42485	A32 - Fidelity World Fund	5	-0.51750	A32 - Fidelity World Fund	4	-0.40300
A42 - SSgA World Index Equity	5	-0.55201	A42 - SSgA World Index Equity	4	-0.42485	A42 - SSgA World Index Equity	5	-0.51750	A42 - SSgA World Index Equity	4	-0.40300
A36 - Franklin Small-Mid Cap Growth	5	-0.55712	A36 - Franklin Small-Mid Cap Growth	4	-0.42996	A36 - Franklin Small-Mid Cap Growth	5	-0.52650	A36 - Franklin Small-Mid Cap Growth	4	-0.41200
A47 - Franklin Global Growth	5	-0.55712	A47 - Franklin Global Growth	4	-0.42996	A47 - Franklin Global Growth	5	-0.52650	A47 - Franklin Global Growth	4	-0.41200
A35 - Transparant B Equity	5	-0.56002	A35 - Transparant B Equity	4	-0.43286	A29 - Pictet European Equity Selection	5	-0.53950	A29 - Pictet European Equity Selection	4	-0.42500
A29 - Pictet European Equity Selection	5	-0.58545	A29 - Pictet European Equity Selection	5	-0.45829	A31 - DWS Top 50 Asia	5	-0.53950	A31 - DWS Top 50 Asia	4	-0.42500
A31 - DWS Top 50 Asia	5	-0.58545	A31 - DWS Top 50 Asia	5	-0.45829	A37 - R Opal Biens Reels	5	-0.53950	A37 - R Opal Biens Reels	4	-0.42500
A37 - R Opal Biens Reels	5	-0.58545	A37 - R Opal Biens Reels	5	-0.45829	A41 - Pictet Europe Index	5	-0.53950	A41 - Pictet Europe Index	4	-0.42500
A41 - Pictet Europe Index	5	-0.58545	A41 - Pictet Europe Index	5	-0.45829	A43 - Pictet Security P	5	-0.53950	A43 - Pictet Security P	4	-0.42500
A43 - Pictet Security P	5	-0.58545	A43 - Pictet Security P	5	-0.45829	A35 - Transparant B Equity	5	-0.54750	A35 - Transparant B Equity	5	-0.43300
A25 - Legg Mason Batterymarch	5	-0.63581	A25 - Legg Mason Batterymarch	5	-0.50865	A25 - Legg Mason Batterymarch	5	-0.57250	A25 - Legg Mason Batterymarch	5	-0.45800
A39 - Vector Navigator C1	5	-0.63581	A39 - Vector Navigator C1	5	-0.50865	A39 - Vector Navigator C1	5	-0.57250	A39 - Vector Navigator C1	5	-0.45800
A44 - GAM Star Global Equity Inflation	5	-0.63581	A44 - GAM Star Global Equity Inflation	5	-0.50865	A44 - GAM Star Global Equity Inflation	5	-0.57250	A44 - GAM Star Global Equity Inflation	5	-0.45800
A48 - Universal Invest Quality Growth	5	-0.63581	A48 - Universal Invest Quality Growth	5	-0.50865	A48 - Universal Invest Quality Growth	5	-0.57250	A48 - Universal Invest Quality Growth	5	-0.45800

For both types of profiles and both experts, the alternatives are sorted into five categories going from "high" social responsibility to "low" social responsibility. The sorting is based on the PROMETHEE II net flows.

Table 6.4: *Sorted funds*

6.4 Application on the Belgian mutual fund market

To illustrate the potential of the proposed sorting tool, we can test for the relative performance of Belgian SRI funds using an unconditional four-factor model (Carhart, 1997) that controls for market risk ($r_{mt} - r_{ft}$), size risk (SMB_t), book-to-market risk (HML_t) and momentum risk (MOM_t).

$$r_{it} - r_{ft} = \alpha_{4i} + \beta_{1i}(r_{mt} - r_{ft}) + \beta_{2i}SMB_t + \beta_{3i}HML_t + \beta_{4i}MOM_t + \varepsilon_{it} \quad (6.1)$$

With r_{it} the equally-weigthed return of mutual fund category i in month t; r_{ft} the risk-free return approximated by a one-month T-bill in month t; $(r_{mt} - r_{ft})$ the excess return on the global market m in month t; SMB_t the return on a zero-investment difference portfolio that is long in a small-cap portfolio and short in a large-cap portfolio; HML_t the return on a zero-investment difference portfolio that is long in a value-type portfolio and short in a growth-type portfolio; and MOM_t the return on a zero-investment difference portfolio that is long in a portfolio with winning stocks and short in a portfolio of losing stocks based on the past year. Assuming that all relevant factors of risk and a representative benchmark have been taken into account, α_{4i} represents the risk-adjusted excess return of mutual fund i. From our new approach towards the definition and classification of mutual funds we overcome a classification bias, and can now test for differences in α_4 between the five different groups of funds to get a better idea about the impact of social responsibility constraints on financial performance of mutual funds.

We retrieve monthly data on our sample of funds from the Morningstar database; risk factor and global market data is gathered from the Kenneth French website, following the methodology of Fama and French (1993) and Carhart (1997). Returns are calculated from monthly price levels quoted in

Euro. The time period for every fund goes back to the month of inception and runs until March of 2014. Results are presented in Table 6.5.

Using a traditional distinction between listed SRI (G1) and conventional funds (G2), following the binary sorting imposed by the financial institution issuing the fund, we find no statistically significant difference in the risk-adjusted excess returns between SRI and conventional funds for the Belgian market. Both types of funds have an alpha that is indistinguishable from zero. Both fund groups have a similar exposure to the global market. The group of conventional funds has a significant negative exposure to the book-to-market factor.

Performing the same analysis on the breakdown of funds in five categories (G1: "most" SRI — G5: "least" SRI), which are built from the process intensity with which the fund was constructed according to ESG factors, we confirm the earlier result for the Belgian case study. Apparently, there are no statistically significant differences in the alpha obtained from the different groups of funds. Every category obtains a zero risk-adjusted excess return. Drawing from a more sophisticated sorting of funds and controlling for classification bias, this finding could serve as a confirmation of earlier results on the performance of SRI versus conventional funds, for the Belgian market. These results are also confirmed using an equally-weighted fund benchmark for every fund category, rather than the global market factor benchmark. From this illustration for the Belgian mutual fund market, we show the potential added value of the FLOWSORT indicator in future mutual fund performance research on larger and more international samples.

6.5 Conclusion

To the best of our knowledge, this is the first application of the FLOWSORT technique in financial economics. From our analysis, we recommend

	Panel A: Estimated coefficients				
			2 groups		
	α_4	MKT	SMB	HML	MOM
G1	-0.067112213	0.783430871	-0.11106925	-0.086649383	-0.039911275
	(0.706854661)	(1.76E-56)	(0.20772861)	(0.328829373)	(0.372225061)
G2	0.086575227	0.779183587	-0.076684484	-0.265136589	-0.007468757
	(0.638809562)	(2.70E-51)	(0.389489113)	(0.00416884)	(0.873984366)
			5 groups		
	α_4	MKT	SMB	HML	MOM
G1	-0.186751956	0.772866611	-0.006777453	-0.02137317	-0.090949022
	(0.327953328)	(6.79E-29)	(0.953436201)	(0.830622603)	(0.097903407)
G2	-0.247954162	0.73734553	-0.132508934	-0.025022748	-0.129931133
	(0.154426623)	(1.35E-57)	(0.195466681)	(0.732280781)	(0.005164332)
G3	-0.237371463	0.744647206	-0.116846648	-0.179966689	-0.063909102
	(0.196670058)	(1.23326E-40)	(0.430837048)	(0.111550397)	(0.190155681)
G4	-0.197225256	0.745742452	-0.166770053	-0.206960597	-0.049474867
	(0.274264923)	(2.96E-48)	(0.126250263)	(0.026425285)	(0.28400555)
G5	-0.102718695	0.724708953	0.075560377	-0.249074029	-0.108427125
	(0.575191038)	(1.60E-40)	(0.528905836)	(0.002577582)	(0.036174483)

	Panel B: Wald tests on α_4			
			2 groups	
	G2			
G1	0.83			
			5 groups	
	G2	G3	G4	G5
G1	0.22	0.27	0.34	0.53
G2	***	0.16	0.2	0.31
G3	***	***	0.24	0.37
G4	***	***	***	0.47

Panel A presents the estimated coefficients from the unconditional four-factor models, using both 2 and 5 groups. P-values for the estimated coefficients are in brackets. Panel B presents the p-values of the Wald tests for significant differences of α_4 between the different groups. Results are obtained from excess fund returns.

Table 6.5: *Unconditional four-factor model results*

that limiting profiles with open classes and five categories are used in future applications. More concretely, the proposed tool can be used in further SRI performance research to introduce more heterogeneity between funds with respect to social responsibility, as was shown in our case study of the Belgian market. One option is to implement multi-factor asset pricing regressions on the five categories of funds, instead of just the group of SRI vs. non-SRI funds. This approach will yield five risk-adjusted returns that can be tested for significant differences in a more nuanced way. The sorting categories can also be used to construct factor-mimicking portfolios to include a so-called ethics risk factor in addition to traditional risk measures (e.g. market risk, size risk, value vs. growth risk and momentum risk). Finally, our tool can be instrumental to assign social responsibility labels to mutual funds, which can be interesting for government regulators looking for curbing the use of the SRI concept for marketing motives.

List of Figures

List of Tables

Bibliography

Alagidede, P. and Panagiotidis, T. (2009). Calendar anomalies in the Ghana Stock Exchange. *Journal of Emerging Market Finance*, 8(1):1–23.

Alexander, S. (1961). Price movements in speculative markets: Trends or random walks. *Industrial Management Review*, 2(2):7–26.

Alexander, S. (1964). Price movements in speculative markets: Trends or random walks, no. 2. *Industrial Management Review*, 5(2):25–46.

Andrews, D. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 58(3):817– 858.

Bachelier, L. (1900). Théorie de la spéculation. *Annales Scientifiques de l'Ecole Normale Supérieure*, 3(17):21–86.

Bana e Costa, C., De Corte, J.-M., and Vansnick, J.-C. (2005). On the mathematical foundation of MACBETH. In Figueira, J., Greco, S., and Ehrgott, M., editors, *Multiple criteria decision analysis: State of the art surveys*, pages 409–437. Springer-Verlag, New York.

Bana e Costa, C. and Vansnick, J. (1994). MACBETH - An interactive path towards the construction of cardinal value functions. *International Transactions in Operational Research*, 1(1):107–114.

Barberis, N., Shleifer, A., and Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics*, 49(3):307–343.

Barnett, M. and Salomon, R. (2006). Beyond dichotomy: The curvilinear relationship between social responsibility and financial performance. *Strategic Management Journal*, 27(11):1101–1122.

Basso, A. and Funari, S. (2003). Measuring the performance of ethical mutual funds: A DEA approach. *Journal of the Operational Research Society*, 54(5):521–531.

Bauer, R., Koedijk, K., and Otten, R. (2005). International evidence on ethical mutual fund performance and investment style. *Journal of Banking & Finance*, 29(7):1751–1767.

Behzadian, M., Kazemzadeh, R., Albadvi, A., and Aghdasi, M. (2010). PROMETHEE: A comprehensive literature review on methodologies and applications. *European Journal of Operational Research*, 200(1):198–215.

Behzadian, M., Otaghsara, K., Yazdani, M., and Ignatius, J. (2012). A state-of-the-art survey of TOPSIS applications. *Expert Systems with Applications*, 39(17):13051–13069.

Belaire-Franch, G. and Contreras, D. (2004). Ranks and signs-based multiple variance ratio tests, University of Valencia working paper.

Belton, V. and Stewart, T. J. (2002). *Multiple criteria decision analysis: An integrated approach.* Kluwer Academic Publishers, Boston.

Bodie, Z., Kane, A., and Marcus, A. (2010). *Investments.* McGraw-Hill Irwin, New York, 9^{th} edition.

Bragdon, J. and Marlin, J. (1972). Is pollution profitable? *Risk Management*, 19(4):9–18.

Brans, J. and Mareschal, B. (1990). The PROMETHEE methods for MCDM: The PROMCALC, GAIA and BANKADVISOR software. In Bana e Costa, C., editor, *Readings in multiple criteria decision aid*, pages 216–252. Springer Verlag, Berlin.

Brans, J. and Vincke, P. (1985). A preference ranking organisation method: The PROMETHEE method for multiple criteria decision-making. *Management Science*, 31(6):647–656.

Brans, J., Vincke, P., and Mareschal, B. (1986). How to rank and how to select projects: The PROMETHEE method. *European Journal of Operational Research*, 24(2):228–238.

Brock, W., Scheinkman, J., Dechert, W., and LeBaron, B. (1996). A test for independence based on the correlation dimension. *Econometric Reviews*, 15(3):197–235.

Brown, D. and Jennings, R. (1989). On technical analysis. *Review of Financial Studies*, 2(4):527–551.

Burashi, A., Sener, E., and Meguturk, M. (2012). The dynamics of limits to arbitrage: Evidence from international cross-sectional data, AFA 2012 Chicago Meetings Paper.

Cabello, J., Ruiz, F., Pérez-Gladish, B., and Méndez-Rodríguez, P. (2014). Synthetic indicators of mutual funds' environmental responsibility: An application of the reference point method. *European Journal of Operational Research*, 236(1):313–325.

Campbell, J., Lo, A., and MacKinlay, A. (1997). *The econometrics of financial markets*. Princeton University Press, Princeton.

Campbell, J. and Yogo, M. (2006). Efficient tests of stock return predictability. *Journal of Financial Economics*, 81(1):27–60.

Carhart, M. (1997). On persistence in mutual fund performance. *Journal of Finance*, 52(1):57–82.

Charles, A. and Darné, O. (2009). Variance ratio tests of random walk: An overview. *Journal of Economic Surveys*, 23(3):503–527.

Charles, A., Darné, O., and Kim, J. (2011). Small sample properties of alternative tests for martingale difference hypothesis. *Economics Letters*, 110(2):151–154.

Checkland, P. and Scholes, J. (1990). *Soft systems methodology in action.* John Wiley & Sons, Chichester.

Chen, W. and Deo, R. (2006). The variance ratio statistic at large horizons. *Econometric Theory*, 22(2):206–234.

Choi, I. (1999). Testing the random walk hypothesis for real exchange rates. *Journal of Applied Econometrics*, 14(3):293–308.

Chow, K. and Denning, K. (1993). A simple multiple variance ratio test. *Journal of Econometrics*, 58(3):385–401.

Christopherson, J., Ferson, W., and Glassman, D. (1998). Conditioning manager alphas on economic information: Another look at the persistence of performance. *Review of Financial Studies*, 11(1):111–142.

Cochrane, J. (2008). The dog that did not bark: A defense of return predictability. *Review of Financial Studies*, 21(4):1533–1575.

Costa, R. and Vasconcelos, G. (2003). Long-range correlations and non-stationarity in the Brazilian stock market. *Physica A*, 329(1-2):231–248.

Cowles, A. (1933). Can stock market forecasters forecast? *Econometrica*, 1(3):309–324.

Cowles, A. (1944). Stock market forecasting. *Econometrica*, 12(3-4):206–214.

Daniel, K., Hirshleifer, D., and Subrahmanyam, A. (1998). Investor psychology and security market under- and overreactions. *Journal of Finance*, 53(6):1839–1885.

Davies, R., Dufour, A., and Scott-Quinn, B. (2006). The MiFID: Competition in a new European equity market regulatory structure. In Ferrarini,

G. and Wymeersch, E., editors, *Investor protection in Europe: Corporate law making, the MiFID and beyond*, chapter 6, pages 161–198. Oxford University Press, Oxford.

De Bondt, W. and Thaler, R. (1985). Does the stock market overreact? *Journal of Finance*, 40(3):793–805.

De Moor, L., Devooght, K., and De Bondt, C. (2012). Een rangschikking volgens het duurzaam en maatschappelijk verantwoord karakter van vier Belgische beleggingsfondsen. *Bank- en Financiewezen*, 76(2-3):161–167.

Delong, B., Shleifer, A., Summers, L., and Waldmann, R. (1990). Positive feedback investment strategies and destabilizing rational expectations. *Journal of Finance*, 45(2):374–397.

Deo, R. and Richardson, M. (2003). On the asymptotic power of the variance ratio test. *Econometric Theory*, 19(2):231–239.

Dimson, E. and Mussavian, M. (1998). A brief history of market efficiency. *European Financial Management*, 4(1):91–103.

Donegan, H., Dodd, F., and McMaster, T. (1992). A new approach to AHP decision-making. *The Statistician*, 41(3):295–302.

Durbin, J. and Koopman, S. (2008). *Time series analysis by state space methods*. Oxford University Press, Oxford.

Engle, R. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007.

Escanciano, J. and Lobato, I. (2009). An automatic Portmanteau test for serial correlation. *Journal of Econometrics*, 151(2):140–149.

Escanciano, J. and Velasco, C. (2006). Generalized spectral tests for the martingale difference hypothesis. *Journal of Econometrics*, 134(1):140–149.

European Fund and Asset Management Association (2013). Asset management in Europe: Facts and figures. Technical report, Brussels.

Eurosif (2010). European SRI study. Technical report, Paris.

Eurosif (2012). European SRI study. Technical report, Brussels.

Fama, E. (1965a). Random walks in stock market prices. *Financial Analysts Journal*, 21(5):55–59.

Fama, E. (1965b). The behavior of stock-market prices. *Journal of Business*, 38(1):34–105.

Fama, E. (1966). Filter rules and stock-market trading. *Journal of Business*, 39(1):226–241.

Fama, E. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, 25(2):383–417.

Fama, E. (1991). Efficient capital markets: II. *Journal of Finance*, 46(5):1575–1617.

Fama, E. (1998). Market efficiency, long-term returns and behavioral finance. *Journal of Financial Economics*, 49(3):283–306.

Fama, E. and French, K. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56.

Fama, E. and French, K. (1996). Multifactor explanations of asset pricing anomalies. *Journal of Finance*, 51(1):55–84.

Febelfin (2012). Aanbeveling inzake duurzame financiële producten. Technical report, Brussels.

Ferson, W. and Schadt, R. (1996). Measuring fund strategy and performance in changing economic conditions. *Journal of Finance*, 51(2):425–461.

Figueira, J., Greco, S., and Ehrgott, M. (2005). *Multiple criteria decision analysis: State of the art surveys*. Springer-Verlag, New York.

Forum for Sustainable and Responsible Investment (2012). Sustainable and responsible investing trends in the US. Technical report, Washington, D.C.

Fox, J. (2009). *The myth of the rational market: A history of risk, reward, and delusion on Wall Street*. Harper Business, New York.

Freeman, R. (1984). *Strategic management: A stakeholder approach*. Pitman/Ballinger, Boston.

Friedman, M. (1970). The social responsibility of business is to increase its profits. *New York Times Magazine*, pp. 32–33, 122, 124, 126.

Geczy, C., Stambaugh, R., and Levin, D. (2003). Investing in socially responsible mutual funds. Working paper.

Gibson, G. (1889). *The stock markets of London, Paris and New York*. G. P. Putnam's Sons, New York.

Granger, C. and Andersen, A. (1978). *An introduction to bilinear time series models*. Vandenhoeck and Ruprecht, Gottingen.

Granger, C. and Morgenstern, O. (1963). Spectral analysis of New York stock market prices. *Kyklos*, 16(1):1–27.

Granovetter, M. (1985). Economic action and social structure: The problem of embeddedness. *American Journal of Sociology*, 91(3):481–510.

Greenwood, R. and Shleifer, A. (2013). Expectations of returns and expected returns, NBER Working Paper No. w18686.

Gromb, D. and Vayanos, D. (2010). The limits of arbitrage: The state of the theory, The Paul Wooley Centre working paper series no. 9 discussion paper no. 650.

Grossman, S. (1976). On the efficiency of competitive stock markets where traders have diverse information. *Journal of Finance*, 31(2):573–585.

Grossman, S. and Stiglitz, J. (1980). On the impossibility of informationally efficient markets. *American Economic Review*, 70(3):393–408.

Gruber, M. (1996). Another puzzle: The growth in actively managed mutual funds. *Journal of Finance*, 51(3):783–810.

Gu, A. and Finnerty, J. (2002). The evolution of market efficiency: 103 years daily data of the Dow. *Review of Quantitative Finance and Accounting*, 18(3):219–237.

Hallerbach, W., Ning, H., Soppe, A., and Spronk, J. (2004). A framework for managing a portfolio of socially responsible investments. *European Journal of Operational Research*, 153(2):517–529.

Hamilton, S., Jo, H., and Statman, M. (1993). Doing well while doing good? The investment performance of socially responsible mutual funds. *Financial Analysts Journal*, 49(6):62–66.

Hammond, J., Keeney, R., and Raiffa, H. (2000). The even-swap method for multiple objective decisions. In Haimes, Y. and Steuer, E., editors, *Research and practice in multiple criteria decision making*, pages 1–14. Springer, Berlin.

Harker, P. and Vargas, L. (1987). The theory of ratio scale estimation: Saaty's analytic hierarchy process. *Management Science*, 33(11):1383–1403.

Hinich, M. (1996). Testing for dependence in the input to a linear time series model. *Journal of Nonparametric Statistics*, 6(2-3):205–221.

Hochberg, Y. (1974). Some conservative generalizations of the T-method in simultaneous inference. *Journal of Multivariate Analysis*, 4(2):224–234.

Hoggett, J. and Nahan, M. (2002). Ethical investment - Deconstructing the myth. *IPA Review*, 54(3):3–6.

Hollingworth, S. (1998). Green investing: A growing concern? *Australian CPA*, 68(4):28–30.

Ishizaka, A. and Labib, A. (2011). Review of the main developments in the analytic hierarchy process. *Expert Systems with Applications*, 38(11):14336–14345.

Ishizaka, A. and Nemery, P. (2013). *Multi-criteria decision analysis: Methods and software*. John Wiley & Sons, New York.

Jacquet-Lagrèze, E. and Siskos, J. (1982). Assessing a set of additive utility functions for multicriteria decision-making: The UTA method. *European Journal of Operational Research*, 10(2):151–164.

Jacquet-Lagrèze, E. and Siskos, J. (1983). *Methode de decision multicritere*. Editions hommes et techniques, Paris.

Jacquet-Lagrèze, E. and Siskos, J. (2001). Preference disaggregation: 20 years of MCDA experience. *European Journal of Operational Research*, 130(2):233–245.

Jain, P. (2005). Financial market design and the equity premium: Electronic versus floor trading. *Journal of Finance*, 60(6):2955–2985.

Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, 48(1):65–91.

Jensen, M. (1968). The performance of mutual funds in the period 1945-1964. *Journal of Finance*, 23(2):389–416.

Jensen, M. (1978). Some anomalous evidence regarding market efficiency. *Journal of Financial Economics*, 6(2-3):95–101.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–292.

Keeney, R. (1992). *Value-focused thinking: A path to creative decision making*. Harvard University Press, Cambridge.

Keeney, R. and Raiffa, H. (1993). *Decisions with multiple objectives: Preferences and value tradeoffs*. John Wiley & Sons, New York.

Keim, D. and Ziemba, W. (2000). Security market imperfections: An overview. In Keim, D. and Ziemba, W., editors, *Security market imperfections in worldwide equity markets*. Cambridge University Press, Cambridge.

Kim, E. and Singal, V. (2000a). Stock market openings: Experience of emerging economies. *Journal of Business*, 73(1):25–66.

Kim, E. and Singal, V. (2000b). The fear of globalizing capital markets. *Emerging Markets Review*, 1(3):183–198.

Kim, J. (2009). Automatic variance ratio test under conditional heteroscedasticity. *Finance Research Letters*, 6(3):179–185.

Kim, J. and Shamsuddin, A. (2008). Are Asian stock markets efficient? Evidence from new multiple variance ratio tests. *Journal of Empirical Finance*, 15(3):518–532.

Kim, J., Shamsuddin, A., and Lim, K. (2011). Stock return predictability and the adaptive markets hypothesis: Evidence from century-long U.S. data. *Journal of Empirical Finance*, 18(5):868–879.

Koellner, T., Weber, O., Fenchel, M., and Scholz, R. (2005). Principles for sustainability rating of investment funds. *Business Strategy and the Environment*, 14(1):54–70.

Kothari, S. (2001). Capital markets research in accounting. *Journal of Accounting and Economics*, 31(1-3):105–231.

KPMG and ALFI (2013). European responsible investing fund survey 2013. Technical report, Luxembourg.

Lakonishok, J., Shleifer, A., and Vishny, R. (1994). Contrarian investment, extrapolation, and risk. *Journal of Finance*, 49(5):1541–1578.

Lambert, R., Leuz, C., and Verrecchia, R. (2012). Information asymmetry, information precision, and the cost of capital. *Review of Finance*, 16(1):1–29.

Lim, K. (2007). Ranking of efficiency for stock markets: A nonlinear perspective. *Physica A*, 376:445–454.

Lim, K. (2008). Sectoral impact of shocks: Empirical evidence from the Malaysian stock market. *Applied Financial Economics Letters*, 4(1):35–39.

Lim, K. and Brooks, R. (2011). The evolution of stock market efficiency over time: A survey of the empirical literature. *Journal of Economic Surveys*, 25(1):69–108.

Lim, K., Luo, W., and Kim, J. (2013). Are US stock index returns predictable? Evidence from automatic autocorrelation-based tests. *Applied Economics*, 45(8):953–962.

Ljung, G. and Box, G. (1978). On a measure of a lack of fit in time series models. *Biometrika*, 65(2):297–303.

Lo, A. (2004). The adaptive markets hypothesis: Market efficiency from an evolutionary perspective. *Journal of Portfolio Management*, 30(5):15–29.

Lo, A. (2005). Reconciling efficient markets with behavioral finance: The adaptive markets hypothesis. *Journal of Investment Consulting*, 7(2):21–44.

Lo, A. and MacKinlay, A. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies*, 1(1):41–66.

Lowenstein, R. (2000). *When genius failed: The rise and fall of long-term capital management*. Random House, New York.

Mackey, A., Mackey, T., and Barney, J. (2007). Corporate social responsibility and firm performance: Investor preferences and corporate strategies. *Academy of Management Review*, 32(3):817–835.

Malkiel, B. (1992). Efficient market hypothesis. In Newman, P., Milgate, M., and Eatwell, J., editors, *New Palgrave dictionary of money and finance*. MacMillan, London.

Malkiel, B. (2003). The efficient market hypothesis and its critics. *Journal of Economic Perspectives*, 17(1):59–82.

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *The Annals of Statistics*, 21(1):255–285.

Mareschal, B. and Brans, J. (1991). BANKADVISER: An industrial evaluation system. *European Journal of Operational Research*, 54(3):318–324.

Mareschal, B., De Smet, Y., and Nemery, P. (2008). Rank reversal in the PROMETHEE II method: Some new results. In *Proceedings of the IEEE International Conference on Industrial Engineering and Engineering Management*, pages 959 – 963, Singapore.

Margolis, J. and Walsh, J. (2003). Misery loves companies: Rethinking social initiatives by business. *Administrative Science Quarterly*, 48(2):268–305.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, 7(1):77–91.

Martel, J., Khoury, N., and Bergeron, M. (1988). An application of a multicriteria approach to portfolio comparisons. *Journal of the Operational Research Society*, 39(7):617–628.

Mitchell, M., Pulvino, T., and Stafford, E. (2002). Limited arbitrage in equity markets. *Journal of Finance*, 57(2):551–584.

Moskowitz, M. (1972). Choosing socially responsible stocks. *Business and Society Review*, 1:71–75.

Narayan, P. and Smyth, R. (2007). Mean reversion versus random walk in G7 stock prices: Evidence from multiple trend break unit root tests. *Journal of International Financial Markets, Institutions and Money*, 17(2):152–166.

Neely, C., Weller, P., and Ulrich, J. (2009). The adaptive markets hypothesis: Evidence from the foreign exchange market. *Journal of Financial and Quantitative Analysis*, 44(2):467–488.

Nemery, P. and Lamboray, C. (2007). FLOWSORT: A flow-based sorting method with limiting or central profiles. *Top*, 16(1):90–113.

Nocera, J. (2009). Poking holes in a theory on markets. *The New York Times*.

Orlitzky, M., Schmidt, F., and Rynes, S. (2003). Corporate social and financial performance: A Meta-analysis. *Organization Studies*, 24(3):403–441.

Patterson, D. and Ashley, R. (2000). *A nonlinear time series workshop: A toolkit for detecting and identifying nonlinear serial dependence*. Kluwer Academic, Boston.

Pérez-Gladish, B. and M'Zali, B. (2010). An AHP-based approach to mutual funds' social performance measurement. *International Journal of Multicriteria Decision Making*, 1(1):103–127.

Petkov, D. and Mihova-Petkova, O. (1997). The analytic hierarchy process and systems thinking. In *13th international MCDM conference*, pages 243–252, Cape Town. Springer.

Petkov, D., Petkova, O., Andrew, T., and Nepal, T. (2007). Mixing multiple criteria decision making with soft systems thinking techniques for decision support in complex situations. *Decision Support Systems*, 43(4):1615–1629.

Phengpis, C. (2006). Are emerging stock market price indices really stationary? *Applied Financial Economics*, 16(13):931–939.

Pöyhönen, M., Hamalainen, R., and Salo, A. (1997). An experiment on the numerical modelling of verbal ratio statements. *Journal of Multi-Criteria Decision Analysis*, 6(1):1–10.

Qu, S., Li, H., and Pei, Y. (2012). Decision making in investing: Application of interval-PROMETHEE based on the composite weight. *Journal of Theoretical and Applied Information Technology*, 45(1):239–245.

Rahman, A. and Saadi, S. (2008). Random walk and breaking trend in financial series: An econometric critique of unit root tests. *Review of Financial Economics*, 17(3):204–212.

Renneboog, L., Ter Horst, J., and Zhang, C. (2008a). Socially responsible investments: Institutional aspects, performance, and investor behavior. *Journal of Banking & Finance*, 32(9):1723–1742.

Renneboog, L., Ter Horst, J., and Zhang, C. (2008b). The price of ethics and stakeholder governance: The performance of socially responsible mutual funds. *Journal of Corporate Finance*, 14(3):302–322.

Renneboog, L., Ter Horst, J., and Zhang, C. (2011). Is ethical money financially smart? Nonfinancial attributes and money flows of socially responsible investment funds. *Journal of Financial Intermediation*, 20(4):562–588.

Roberts, H. (1967). Statistical versus clinical prediction of the stock market. Technical report.

Ross, S. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3):341–360.

Roy, B. (1974). Critères multiples et modélisation des préférences: L'apport des relations de surclassement. *Revue d'Economie Politique*, 1:1–44.

Roy, B. (1985). *Méthodologie multicritère d'aide à la décision*. Economica, Paris.

Roy, B. (1996). *Multicriteria methodology for decision analyis*. Kluwer Academic Publishers, Dordrecht.

Saaty, T. (1977). A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*, 15(3):234–281.

Saaty, T. (1980). Analytic hierarchy process. *Decision Analysis*, 50(1):579–606.

Saaty, T. (2005). The analytic hierarchy and analytic network processes for the measurement of intangible criteria and for decision-making. In Figueira, J., Greco, S., and Ehrgott, M., editors, *Multiple criteria decision analysis: State of the art surveys*, pages 345–408. Springer.

Saaty, T. and Takizawa, M. (1986). Dependence and independence: From linear hierarchies to nonlinear networks. *European Journal of Operational Research*, 26(2):229–237.

Samuelson, P. (1965). Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review*, 6(2):41–49.

Sauer, D. (1997). The impact of social-responsibility screens on investment performance: Evidence from the Domini 400 social index and Domini equity mutual fund. *Review of Financial Economics*, 6(2):137–149.

Schwartz, M. (2003). The "ethics" of ethical investing. *Journal of Business Ethics*, 43(3):195–213.

Schwert, G. (2003). Anomalies and market efficiency. In Constantinides, G., Harris, M., and Stulz, R., editors, *Handbook of the economics of finance*. Elsevier B.V., Amsterdam.

Sewell, M. (2011). History of the efficient market hypothesis, University College of London Working paper RN/11/04.

Sharpe, W. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19(3):425–442.

Shefrin, H. (2000). *Beyond greed and fear: Understanding behavioral finance and the psychology of investing*. Oxford University Press, Oxford.

Shiller, R. (2000). *Irrational exuberance*. Princeton University Press, Princeton.

Shiller, R. (2003). From efficient markets theory to behavioral finance. *Journal of Economic Perspectives*, 17(1):83–104.

Shleifer, A. (2000). *Inefficient markets: An introduction to behavioral finance*. Oxford University Press, Oxford.

Shleifer, A. and Vishny, R. (1997). The limits of arbitrage. *Journal of Finance*, 52(1):35–55.

Simon, H. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, 69(1):99–118.

Smith, A. (1759). *The theory of moral sentiments*. Liberty Fund, Indianapolis.

Smith, A. (1766). The wealth of nations. In *The wealth of nations*. P.F. Collier, New York.

Sparkes, R. and Cowton, C. (2004). The maturing of socially responsible investment: A review of the developing link with corporate social responsibility. *Journal of Business Ethics*, 52(1):45–57.

Steuer, R. and Na, P. (2003). Multiple criteria decision making combined with finance: A categorized bibliographic study. *European Journal of Operational Research*, 150(3):496–515.

Tabak, B. (2003). The random walk hypothesis and the behaviour of foreign capital portfolio flows: The Brazilian stock market case. *Applied Financial Economics*, 13(5):369–378.

Thaler, R. (1992). *The winner's curse: Paradoxes and anomalies of economic life*. Princeton University Press, Princeton.

Tippet, J. (2001). Performance of Australia's ethical funds. *Australian Economic Review*, 34(2):170–178.

Todea, A., Ulici, M., and Silaghi, S. (2009). Adaptive markets hypothesis: Evidence from Asia-Pacific financial markets. *Review of Finance and Banking*, 1(1):7–13.

Tsao, C. (2003). Evaluating investment values of stocks using a fuzzy TOPSIS approach. *Journal of Information and Optimization Sciences*, 24(2):373–396.

United Nations (2013). The six principles for responsible investment. Technical report.

Urquhart, A. and Hudson, R. (2013). Efficient or adaptive markets? Evidence from major stock markets using very long run historic data. *International Review of Financial Analysis*, 28:130–142.

Van den Bossche, F., Rogge, N., Devooght, K., and Van Puyenbroeck, T. (2010). Robust corporate social responsibility investment screening. *Ecological Economics*, 69(5):1159–1169.

Verheyden, T. (2013). A tale of market efficiency: A methodological digress. Master thesis KU Leuven (Brussels Campus).

Verheyden, T. and De Moor, L. (2014). Process-oriented social responsibility indicator for mutual funds: A multi-criteria decision analysis approach. Working paper. Available at SSRN: http://ssrn.com/abstract=2386566.

Verheyden, T., Van den Bossche, F., and De Moor, L. (2014). Towards a new framework on efficient markets. Working Paper. Available at SSRN: http://ssrn.com/abstract=2382809.

von Neumann, J. and Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton University Press, Princeton.

Waddock, S. and Graves, S. (1997). The corporate social performance - financial performance link. *Strategic Management Journal*, 18(4):303–319.

Wallenius, J., Dyer, J., Fishburn, P., Steuer, R., Zionts, S., and Deb, K. (2008). Multiple criteria decision making, multiattribute utility theory: Recent accomplishments and what lies ahead. *Management Science*, 54(7):1336–1349.

Wang, X. and Triantaphyllou, E. (2008). Ranking irregularities when evaluating alternatives by using some ELECTRE methods. *Omega*, 36(1):45–63.

Wermers, R. (2000). Mutual fund performance: An empirical decomposition into stock-picking talent, style, transaction costs, and expenses. *Journal of Finance*, 55(4):1655–1703.

Working, H. (1934). A random-difference series for use in the analysis of time series. *Journal of the American Statistical Association*, 83(1):87–93.

Wright, J. (2000). Alternative variance-ratio tests using ranks and signs. *Journal of Business & Economic Statistics*, 18(1):1–9.

Zalewska-Mitura, A. and Hall, S. (1999). Examining the first stages of market performance. A test for evolving market efficiency. *Economics Letters*, 64(1):1–12.

Zhou, J. and Lee, J. (2013). Adaptive market hypothesis: Evidence from the REIT market. *Applied Financial Economics*, 23(21):1649–1662.

Zopounidis, C. and Doumpos, M. (1999). Business failure prediction using the UTADIS multicriteria analysis method. *Journal of the Operational Research Society*, 50(11):1138–1148.

Zopounidis, C. and Doumpos, M. (2002). Multicriteria classification and sorting methods: A literature review. *European Journal of Operational Research*, 138(2):229–246.

Doctoral dissertations from the Faculty of Business and Economics

A list of doctoral dissertations from the Faculty of Business and Economics can be found at the following website:
`http://www.kuleuven.be/doctoraatsverdediging/archief.htm`.